Département de géomatique appliquée
Faculté des lettres et sciences humaines
Université de Sherbrooke

Modélisation tridimensionnelle précise de l'environnement à l'aide des systèmes de photogrammétrie embarqués sur drones

On precise three-dimensional environment modeling via UAV-based photogrammetric systems

Mozhdeh Shahbazi

Directeur de recherche: Jérome Théau
Co-directeur de recherche: Gunho Sohn

Thèse présentée pour l'obtention du grade Philosophiae Doctor (PhD) en télédétection
Cheminement en géomatique appliquée

juillet 2016

Composition du jury

Modélisation tridimensionnelle précise de l'environnement à l'aide des systèmes de photogrammétrie embarqués sur drones

On precise three-dimensional environment modeling via UAV-based photogrammetric systems

Mozhdeh Shahbazi

Cette thèse a été évaluée par un jury composé des personnes suivantes:

Professor Dr. Jérome Théau
Dép. de géomatique appliquée, Université de Sherbrooke, Sherbrooke, QC, Canada
Directeur de recherche

Professor Dr. Gunho Sohn
Dep. of Geomatics Engineering, York University, Toronto, ON, Canada
Co-directeur de recherche

Professor Dr. Ayman Habib
Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, USA
Examinateur externe

Professor Dr. Kalifa Goïta
Dép. de géomatique appliquée, Université de Sherbrooke, Sherbrooke, QC, Canada
Examinateur interne

Professor Dr. Jean-Marie Dubois
Dép. de géomatique appliquée, Université de Sherbrooke, Sherbrooke, QC, Canada
Examinateur interne

Abstract

Images acquired from unmanned aerial vehicles (UAVs) can provide data with unprecedented spatial and temporal resolution for three-dimensional (3D) modeling. Solutions developed for this purpose are mainly operating based on photogrammetry concepts, namely UAV-Photogrammetry Systems (UAV-PS). Such systems are used in applications where both geospatial and visual information of the environment is required. These applications include, but are not limited to, natural resource management such as precision agriculture, military and police-related services such as traffic-law enforcement, precision engineering such as infrastructure inspection, and health services such as epidemic emergency management.

UAV-photogrammetry systems can be differentiated based on their spatial characteristics in terms of accuracy and resolution. That is some applications, such as precision engineering, require high-resolution and high-accuracy information of the environment (e.g. 3D modeling with less than one centimeter accuracy and resolution). In other applications, lower levels of accuracy might be sufficient, (e.g. wildlife management needing few decimeters of resolution). However, even in those applications, the specific characteristics of UAV-PSs should be well considered in the steps of both system development and application in order to yield satisfying results.

In this regard, this thesis presents a comprehensive review of the applications of unmanned aerial imagery, where the objective was to determine the challenges that remote-sensing applications of UAV systems currently face. This review also allowed recognizing the specific characteristics and requirements of UAV-PSs, which are mostly ignored or not thoroughly assessed in recent studies.

Accordingly, the focus of the first part of this thesis is on exploring the methodological and experimental aspects of implementing a UAV-PS. The developed system was extensively evaluated for precise modeling of an open-pit gravel mine and performing volumetric-change measurements. This application was selected for two main reasons. Firstly, this case study provided a challenging environment for 3D modeling, in terms of scale changes, terrain relief variations as well as structure and texture diversities. Secondly, open-pit-mine monitoring demands high levels of accuracy, which justifies our efforts to improve the developed UAV-PS to its maximum capacities. The hardware of the system consisted of an electric-powered helicopter, a high-resolution digital camera, and an inertial navigation system. The software of the system included the in-house programs specifically designed for camera calibration, platform calibration, system integration, onboard data acquisition, flight planning and ground control point (GCP) detection. The detailed features of the system are discussed in the thesis, and solutions are proposed in order to enhance the system and its photogrammetric outputs. The accuracy of the results was evaluated under various mapping conditions, including direct georeferencing and indirect georeferencing with different numbers, distributions and types of ground control points. Additionally, the effects of imaging configuration and network stability on modeling accuracy were assessed.

The second part of this thesis concentrates on improving the techniques of sparse and dense reconstruction. The proposed solutions are alternatives to traditional aerial photogrammetry techniques, properly adapted to specific characteristics of unmanned, low-altitude imagery. Firstly, a method was

developed for robust sparse matching and epipolar-geometry estimation. The main achievement of this method was its capacity to handle a very high percentage of outliers (errors among corresponding points) with remarkable computational efficiency (compared to the state-of-the-art techniques). Secondly, a block bundle adjustment (BBA) strategy was proposed based on the integration of intrinsic camera calibration parameters as pseudo-observations to Gauss-Helmert model. The principal advantage of this strategy was controlling the adverse effect of unstable imaging networks and noisy image observations on the accuracy of self-calibration. The sparse implementation of this strategy was also performed, which allowed its application to data sets containing a lot of tie points. Finally, the concepts of intrinsic curves were revisited for dense stereo matching. The proposed technique could achieve a high level of accuracy and efficiency by searching only through a small fraction of the whole disparity search space as well as internally handling occlusions and matching ambiguities. These photogrammetric solutions were extensively tested using synthetic data, close-range images and the images acquired from the gravel-pit mine. Achieving absolute 3D mapping accuracy of 11±7 mm illustrated the success of this system for high-precision modeling of the environment.

Résumé

Les images acquises à l'aide d'aéronefs sans pilote (ASP) permettent de produire des données de résolutions spatiales et temporelles uniques pour la modélisation tridimensionnelle (3D). Les solutions développées pour ce secteur d'activité sont principalement basées sur des concepts de photogrammétrie et peuvent être identifiées comme des systèmes photogrammétriques embarqués sur aéronefs sans pilote (SP-ASP). Ils sont utilisés dans plusieurs applications environnementales où l'information géospatiale et visuelle est essentielle. Ces applications incluent notamment la gestion des ressources naturelles (ex. : agriculture de précision), la sécurité publique et militaire (ex. : gestion du trafic), les services d'ingénierie (ex. : inspection de bâtiments) et les services de santé publique (ex. : épidémiologie et gestion des risques).

Les SP-ASP peuvent être subdivisés en catégories selon les besoins en termes de précision et de résolution. En effet, dans certains cas, tel qu'en ingénierie, l'information sur l'environnement doit être de haute précision et de haute résolution (ex. : modélisation 3D avec une précision et une résolution inférieure à un centimètre). Pour d'autres applications, tel qu'en gestion de la faune sauvage, des niveaux de précision et de résolution moindres peut être suffisants (ex. : résolution de l'ordre de quelques décimètres). Cependant, même dans ce type d'applications les caractéristiques des SP-ASP devraient être prises en considération dans le développement des systèmes et dans leur utilisation, et ce, pour atteindre les résultats visés.

À cet égard, cette thèse présente une revue exhaustive des applications de l'imagerie aérienne acquise par ASP et de déterminer les challenges les plus courants. Cette étude a également permis d'établir les caractéristiques et exigences spécifiques des SP-ASP qui sont généralement ignorées ou partiellement discutées dans les études récentes.

En conséquence, la première partie de cette thèse traite des aspects méthodologiques et d'expérimentation de la mise en place d'un SP-ASP. Le système développé a été évalué pour la modélisation précise d'une gravière et utilisé pour réaliser des mesures de changement volumétrique. Cette application a été retenue pour deux raisons principales. Premièrement, ce type de milieu fournit un environnement difficile pour la modélisation, et ce, en termes de changement d'échelle, de changement de relief du terrain ainsi que la grande diversité de structures et de textures. Deuxièment, le suivi de mines à ciel ouvert exige un niveau de précision élevé, ce qui justifie les efforts déployés pour mettre au point un SP-ASP de haute précision. Les composantes matérielles du système consistent en un ASP à propulsion électrique de type hélicoptère, d'une caméra numérique à haute résolution ainsi qu'une station inertielle. La composante logicielle est composée de plusieurs programmes développés particulièrement pour calibrer la caméra et la plateforme, intégrer les systèmes, enregistrer les données, planifier les paramètres de vol et détecter automatiquement les points de contrôle au sol. Les détails complets du système sont abordés dans la thèse et des solutions sont proposées afin d'améliorer le système et la qualité des données photogrammétriques produites. La précision des résultats a été évaluée sous diverses conditions de cartographie, incluant le géoréférencement direct et indirect avec un nombre, une répartition et des types de points de contrôle variés. De plus, les effets de la configuration des images et la stabilité du réseau sur la précision de la modélisation ont été évalués.

La deuxième partie de la thèse porte sur l'amélioration des techniques de reconstruction éparse et dense. Les solutions proposées sont des alternatives aux techniques de photogrammétrie aérienne traditionnelle et adaptée aux caractéristiques particulières de l'imagerie acquise à basse altitude par ASP. Tout d'abord, une méthode robuste de correspondance éparse et d'estimation de la géométrie épipolaire a été développée. L'élément clé de cette méthode est sa capacité à gérer le pourcentage très élevé des valeurs aberrantes (erreurs entre les points correspondants) avec une efficacité de calcul remarquable en comparaison avec les techniques usuelles.

Ensuite, une stratégie d'ajustement de bloc basée sur l'intégration de pseudoobservations du modèle Gauss-Helmert a été proposée. Le principal avantage de cette stratégie consistait à contrôler les effets négatifs du réseau d'images instable et des images bruitées sur la précision de l'autocalibration. Une implémentation éparse de cette stratégie a aussi été réalisée, ce qui a permis de traiter des jeux de données contenant des millions de points de liaison. Finalement, les concepts de courbes intrinsèques ont été revisités pour l'appariement stéréo dense. La technique proposée pourrait atteindre un haut niveau de précision et d'efficacité en recherchant uniquement dans une petite portion de l'espace de recherche des disparités ainsi qu'en traitant les occlusions et les ambigüités d'appariement. Ces solutions photogrammétriques ont été largement testées à l'aide de données synthétiques, d'images à courte portée ainsi que celles acquises sur le site de la gravière. Le système a démontré sa capacité a modélisation dense de l'environnement avec une très haute exactitude en atteignant une précision 3D absolue de l'ordre de 11±7 mm.

Acknowledgements

I would like to thank Dr. Jérome Théau for his supervision, his patience at my difficult moments, his trust on me and his support to initiate my academic life in Canada. I would like to acknowledge his great knowledge of applied remote sensing, which helped me to achieve my research objectives with this regard. Starting my Ph.D. career at Université de Sherbrooke has been due to his belief in my capacities, which I appreciate enormously.

I would like to express my sincere gratitude to my co-supervisor Dr. Gunho Sohn for his generous support of my work, his patience and inspiration, his immense knowledge of photogrammetry, and his endless source of ideas. I would like to thank him for giving me the opportunity of working at York University and joining brainstorming sessions in his lab, and for introducing me to his great international, industrial and academic network. I could not have imagined having a better mentor in my academic life.

This thesis was supported in part by grants from Centre de Géomatique du Québec (CGQ), Fonds de Recherche Québécois sur la Nature et les Technologies (FRQNT), and Natural Sciences and Engineering Research Council of Canada (NSERC). Many thanks to Mr. Patrick Ménard, my industrial advisor at CGQ, without whose generous technical supports for system development and his assistance for data acquisition this thesis would not be accomplished. I would like to thank Mrs. Josée Dalaire, the CEO of CGQ, for her kind and encouraging support of my research even after finishing the period of my scholarship.

I would like to acknowledge and thank Dr. Wolfgang Förstner for his great ideas to improve the saprse matching article, his generous devotion of his time to revise the manuscript and sharing his incredible knowledge of computational stereo with me.

I would like to thank the rest of my thesis committee: Dr. Ayman Habib, Dr. Kalifa Goïta, and Dr. Jean-Marie Dubois, for agreeing to review this dissertation, for their insightful and constructive comments as well as their encouragements. It has been my honor to receive their comments.

The gravel-pit data sets were acquired at Bel-horizon site belonging to DJL Construction Inc. I would like to thank Mr. Michel Laperle, the manager of the site, for his kind support to facilitate the fieldwork. I would like to thank Mr. Kim Desrosiers, Mr. Erwan Gavelle, and ING Robotics' team members for their assistance in the field work.

I would like to thank Mr. Michael Cohen, the president of Industrial SkyWorks, for providing me with the opportunity of collaborating with his company in a research project, without the financial support of which I could not concentrate on my thesis.

I'm thankful to my office-mates at USherbrooke for their friendly support of me as a non-francophone member, to my lab mates at YorkU for sharing their research experiences with me, and to my colleagues at CGQ for welcoming me in Saguenay region.

Finally yet importantly, I am grateful for the support, kindness and patience of my family during the last few years, even though I could not take a good care of them due to my overloaded volume of work.

Table of Contents

List of Figures

List of Tables

Acronyms

1D	One Dimensional
2D	Two Dimensional
3D	Three Dimensional
AD	Absolute Difference
BA	Bundle Adjustment
BBA	Block Bundle Adjustment
BP	Belief Propagation
CCD	Charge Coupled Device
DLT	Direct Linear Transformation
DoF	Depth of Field
DP	Dynamic Programming
DSI	Disparity Space Image
DSM	Digital Surface Model
DTM	Digital Terrain Model
ECEF	Earth-Centered, Earth-Fixed
EO	Exterior Orientation
FoV	Field of View
GA	Genetic Algorithm
GC	Graph Cut
GCP	Ground Control Point
GFoV	GroundField of View
GHM	Gauss-Helmert Model
GMM	Gauss-Markov Model
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
iBA	Incremental Bundle Adjustment
IC	Intrinsic Curves
IMU	Inertial Measurement Unit

INS	Inertial Navigation System
IO	Interior Orientation
LBP	Loopy Belief Propagation
LPS	Local Plane Sweeps
LM	Levenberg–Marquardt
LU	Lower Upper
MRF	Markov Random Fields
NCC	Normalized Cross Correlation
NP-hard	Non-deterministic Polynomial-time hard
PA	Precision Agriculture
PCG	Preconditioned Conjugate Gradients
RANSAC	Random Sample Consensus
RMS	Root Mean Square
RO	Relative Orientation
RS	Remote Sensing
RTK	Real-Time Kinematics
SfM	Structure from Motion
SGM	Semi-Global Matching
SIFT	Scale-Invariant Feature Transform
SLAM	Simultaneous Localization and Mapping
SSD	Sum of Squared Differences
StD	Standard Deviation
SVD	Single Value Decomposition
SVM	Support Vector Machine
TV	Total Variance
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle
UAV-PS	UAV-Photogrammetry System
WAAS	Wide Area Augmentation System

Reading Guide

As part of this research project, we have published two articles in scientific ISI journals, GIScience and Remote Sensing as well as Sensors. Two articles are also prepared for submission to Computer Vision and Image Understanding as well as ISPRS Journal of Photogrammetry and Remote Sensing.

Therefore, this thesis is presented in the form of articles based on our four journal publications/submissions. Four technical chapters of the thesis are devoted to the articles. Each chapter contains the following sections: background, general methodology and the original text of the article. Chapter 1 is an exception, which contains the general research background, objectives, and hypotheses. Chapter 5 is also an exception, where extensions are added to Chapter 4 to complete the process of sparse reconstruction with a focus on the strategy of block bundle adjustment. Finally, Chapter 7 is dedicated to overall discussions, conclusions, and future work.

Chapter 1

1. Introduction

1.1 Unmanned Aerial Photogrammetry

Photogrammetry is defined as a set of metrology techniques for three-dimensional measurement of the objects and phenomena indirectly from overlapped images captured from at least two different orientations, namely stereoscopic viewing (Linder, 2006). With this definition, three physical elements are involved in a photogrammetric system: the imaging device (camera), the platform carrying it and the study object. Accordingly, photogrammetry fields can be divided into close-range and aerial ones. In close-range photogrammetry, the distance between the object and the camera is closer than few meters, and the imaging platform is either a handheld device or a terrestrial vehicle (Luhmann et al., 2007). In contrary, aerial photogrammetry is involved with airborne platforms and large (hundreds of meters) distances from the objects. Recently, the advancement of unmanned aerial vehicles (UAVs) and improvement of high-resolution consumer-grade cameras has initiated a new field, namely UAV-photogrammetry.

In general, the techniques of photogrammetry can be divided into two categories. The first group performs conventional photogrammetry on the whole block of acquired images. The techniques of the second group are based on sequential structure-from-motion (SfM) computation, which is performed by successive reconstruction. The second category is applied when real-time mapping from video frames is required. In this thesis, the focus is on conventional photogrammetry. However, the application of the developed algorithms and their efficiency for sequential SfM is also discussed, whenever applicable.

In either of the techniques, the main steps of photogrammetric processing include the followings:

- sparse image matching and epipolar-geometry[1] estimation,
- block bundle adjustment,
- dense image matching performed on rectified images, and
- triangulation and 3D point cloud generation.

In computer-vision vocabulary, the term *Structure from Motion* is used for referring to the process of sparse matching and *motion* estimation, as a result of which the sparse 3D *structure* of the scene can be reconstructed. Similarly, the term *dense reconstruction* is used for referring to the process of dense matching and 3D point cloud generation. The following paragraphs briefly explain the general

[1] Epipolar geometry is the general term used to describe the geometry of stereo vision. There are a number of geometric constraints between the 3D points and their projections onto the 2D images, e.g. epipoles, epipolar lines, epipolar planes. In this thesis, the epipolar constraints are described by the essential matrix or the fundamental matrix between the two cameras, depending on the availability of intrinsic camera calibration parameters.

description of these steps. However, thorough introduction to the state-of-the-art techniques as well as their challenges and issues can be found in the next chapters of the thesis.

Given two overlapping images, sparse corresponding points (known as tie points, conjugate points, or matches) are detected. These points are selected from salient areas of the image, where reliable, unambiguous matching can be performed. These salient areas are represented by features, which can be points, lines or regions. Different types of descriptors are used to describe the features (Mukherjee et al., 2015; Li and Allinson, 2008; Apollonio et al., 2014). The similarity between the descriptors of one feature in the first image and those of another feature in the second image determines whether they are corresponding or not. These tie points (as image observations) are used to estimate the initial values of relative orientation[1] parameters and 3D coordinates of the points.

In the case of conventional photogrammetry, image observations of a complete block of images and the initial estimates of 3D object coordinates and orientation parameters are adjusted through block bundle adjustment (BBA). Often, the intrinsic camera(s) calibration parameters are also estimated in the BBA as additional parameters (Ackermann, 1981). The principal outputs of BBA are refined exterior orientation (EO) parameters of images, intrinsic camera calibration parameters, and sparse structure. In the case of sequential SfM, incremental bundle adjustment (iBA) is performed on the observations of some of the frames, called key frames. This aims at avoiding the accumulation of errors in large-scale image sequences while achieving high computational efficiency (Indelman et al., 2012).

Exterior orientation parameters are used to facilitate dense reconstruction since they add constraints to the relative position of corresponding points. The most common way to apply these constraints is to rectify[2] images first and to perform dense stereo/multi-view matching on rectified images then. That is the corresponding points are located on the same horizontal lines (scanlines) of images. These constraints can also be added explicitly in the objective function of matching, as in dense simultaneous localization and mapping (SLAM) (Engel et al., 2014). The result of dense matching is the disparity map, which represents the offsets (x-parallax) between matches. Disparity maps can be transformed to depth and 3D coordinates using the EO parameters by a process called intersection[3]. Georeferencing the 3D point cloud can be performed along with the BBA using the known coordinates of the ground control points (GCPs), called indirect geo-referencing. It can also be done directly using the exterior orientation parameters of the images measured by the navigation sensors, known as direct geo-referencing. Geo-referencing may also be performed in a separate step from BBA by finding a simple 3D transformation between the arbitrary coordinate system of the generated point cloud and the world coordinate system.

---

[1] Relative orientation parameters are the parameters of a rigid-body transformation that are required to define the relationship between the photo-coordinates of a stereo-pair.

[2] Rectification is the process of making stereo image planes parallel to their baseline by defining two new perspective projection matrices that preserve the optical centres. This ensures that epipoles are moved to infinity, so the epipolar lines become parallel. Moreover, the epipolar line of any point $p=(x_2,y_2)$ in the right image becomes a horizontal line as $y_2=y_1$ in the left image, with $p'=(x_1,y_1)$ being the conjugate of p (Fusiello et al., 2000).

[3] In vocabulary of computer vision, the term triangulation might be used (Granshaw and Fraser, 2015).

1.2 Challenges of UAV-Photogrammetry

Surveying-grade UAV-photogrammetry systems (UAV-PSs) have significant differences from traditional photogrammetry systems, which should be considered carefully in their development and applications. Nowadays, different pre-packaged systems are available commercially, where the sensors and their characteristics are pre-selected. These systems offer certain operational advantages, such as safety and technical support of the manufacturer. However, they offer little flexibility regarding sensor selection and adjustment. When specific requirements of an application cannot be met by such systems, the best practice is developing a system by integrating individual equipment, which best fits the application. In Chapter 3, the research problems with this regard are discussed in details.

The photogrammetric solutions adapted to UAV-PSs should be able to deal with specific characteristics of low-altitude imagery. The most distinctive characteristics of this type of imagery include the followings:

i) Geometric characteristics: high resolution, large perspective distortion, and scale changes due to oblique photography and low flight altitude in comparison with terrain relief, and
ii) Photometric characteristics: uneven distribution of key-points, motion blur, out-of-focus blur, occlusion, foreground motion of features, and noticeable radiometric changes (Zhang et al., 2011; Haala et al., 2013; Hu and Ai, 2011).

Generally, UAV-PSs result in large amounts of high-resolution images with the mentioned characteristics. Therefore, the sparse matching task becomes more challenging and the percentage of outliers (false matches) grows highly. For instance, less outlier would result by matching tree crowns at low-resolution images compared with matching tree leaves in high-resolution images. Hence, robust algorithms with high computational efficiency are essentially required to remove these outliers. Moreover, they need to be robust against image noise, which is a common problem in UAV imagery due to platform characteristics.

Similarly, the techniques of dense image matching should be computation-effective to deal with large volumes of high-resolution images. These techniques should also be robust against irregular range disparities and presence of occlusions, which are common effects in low-altitude imagery.

To the best of the author's knowledge, not all the issues and concerns mentioned herein are considered simultaneously in a comprehensive project for development of a UAV-photogrammetry system.

1.3 Objectives and Assumptions

The general objective of this thesis is to develop a UAV-photogrammetry system that can respond to the 3D-modeling requirements of various engineering and natural-resource-management applications, in terms of accuracy, spatial resolution, and temporal operability. The followings are the particular objectives of this thesis.

1. Studying the applications of unmanned aerial imagery

- Hypothesis: Understanding pros, cones, issues and technical challenges of recent UAV-applications, as well as the shortcomings of previously developed systems, would allow recognizing the requirements of a high-performance system.

2. Exploring the methodological and experimental aspects of implementing a UAV-PS for ensuring high-quality visual and topographic data

- Hypothesis: A UAV-PS can produce frequent, high-resolution and accurate topographic data if its capacities are maximized by considering the following elements.
  - Careful camera calibration and assessment of its metric characteristics using a specific test-field configuration, automatic target detection, and stability analysis
  - Robust system integration to control and synchronize the navigation and imaging sensors accurately (up to the restrictions of the sensors) without needing any additional mechanism
  - Pre-planning the flight and data acquisition scenarios
  - Using a minimal number of GCPs by optimizing their distribution, maximizing their visibility, ensuring the scale consistency of the network, and accurate localization of GCPs on images
  - Enhancing the radiometric quality of acquired images by pre-processing

These hypotheses are evaluated in the specific application of surveying and volumetric change measurement in an open-pit gravel mine.

3. Developing photogrammetric solutions that are adapted to the characteristics of low-altitude imagery. The specific objectives with this regard include the followings.

3.1. Developing a technique of robust epipolar-geometry estimation and sparse matching based on evolutionary search

- Hypothesis: Conventional random consensus sampling techniques of robust estimation are insufficient to efficiently deal with a very high percentage of errors while the proposed technique is robust to both high outlier and noise ratios.

3.2. Assessing a BBA strategy for on-the-job self-calibration based on integration of pseudo-observations to Gauss-Helmert model

- Hypothesis: The specific configuration of UAV-imaging networks causes high correlation between interior and exterior orientation parameters of cameras, which adversely affect the accurate estimation of intrinsic camera calibration parameters. The proposed strategy can numerically de-correlate the unknowns to increase the accuracy of BBA and independent 3D reconstruction.

3.3. Revisiting the concepts of intrinsic curves for developing an efficient dense stereo matching technique

- Hypothesis: The proposed technique can provide high efficiency and accuracy by searching only through a small fraction of the whole disparity search space (using the original resolution of images) and internally handling occlusions and matching ambiguities.

These hypotheses are mainly evaluated using close-range images since it is believed that the characteristics of low-altitude aerial imagery are very similar to close-range imagery. They are additionally tested by open-pit mine datasets.

1.4 Thesis Structure

This dissertation is organized into six chapters. In the current chapter, the general context and research problematic were presented. Moreover, the objectives and assumptions of the research were defined. The rest of the thesis is organized as follows. The subsequent chapters are devoted to four scientific papers that were written in the course of this research.

Accordingly, Chapter 2 presents our review study on the applications of unmanned aerial imagery. This chapter is formed of an article entitled “Recent Applications of Unmanned Aerial Imagery in Natural Resource Management” which is published in Volume 51 of GIScience & Remote Sensing Journal.

Chapter 3 presents different aspects of system development, data acquisition and application in open-pit mine surveying. This chapter is formed of an article entitled “Development and Evaluation of a UAV-Photogrammetry System for Precise 3D Environmental Modeling” which is published in Volume 15 of Sensors journal.

Chapter 4 presents the first part of the proposed technique for sparse reconstruction, including robust sparse matching and epipolar-geometry estimation. This chapter is formed of a manuscript entitled “Evolutionary Optimization for Robust Sparse Matching and Epipolar-Geometry Estimation” which is prepared for submission to Computer Vision and Image Understanding journal.

Chapter 5 presents the second part of the proposed technique for sparse reconstruction, including the structure estimation and bundle adjustment. This chapter, due to its short length, is not presented in the format of an article.

Chapter 6 presents the proposed technique of dense reconstruction, with a focus on dense stereo matching. This chapter is formed of a manuscript entitled “A Dense Stereo Matching Method Using Intrinsic Curves” which is prepared for submission to ISPRS journal of Photogrammetry and Remote Sensing.

Chapter 7 is dedicated to the overall conclusions, the recommendations, and the future work.

Chapter 2

2. Recent Applications of Unmanned Aerial Imagery

2.1 Article Presentation

2.1.1 Background

This article is related to the first specific objective of the thesis, namely studying the recent application of unmanned aerial imagery. This article was originally published in GIScience & Remote Sensing Aug 2014 © Informa UK Limited, trading as Taylor & Francis Group, available online: 10.1080/ 15481603.2014.926650.

Currently, there are hundreds of commercial developers, whose focus is on the hardware development of unmanned aerial systems. However, less attention has been paid to the development and evaluation of data processing techniques and system requirements with regard to specific applications. Therefore, critical reviews of previous studies are required to assess the current technological and scientific status of unmanned remote sensing (RS) applications.

In this study, a specific category of RS applications was considered, namely natural resource management via unmanned aerial passive imagery. The first reason for this choice was the vast range of applications that could be classified under the flag of natural resource management, such as precision agriculture, rangeland monitoring, natural disaster management, aquatic ecosystem management, polar remote sensing and wildlife research. Therefore, this review could give us a broad insight into unmanned aerial systems dedicated to remote sensing and photogrammetry. The second reason was that all these applications had one technical element in common: they all required visual and geospatial information of the environment. This feature allowed us to study their requirements in terms of imaging sensors and navigation systems. The conclusions made at the end of this study allowed us to make clear decisions regarding the components of our system. The last reason was that surveying and monitoring of open-pit mines (the case study of this thesis) could be categorized in this series of applications.

2.1.2 General methodology

Although more than 200 articles were studied to conduct this review, it was decided to narrow down the research by only considering the articles that represented a complete/semi-complete data-processing chain to achieve their objectives. This selection was made to prove a point, that image acquisition with a UAV is no more a challenging topic in either industry or academia.

In the course of this article, each application is divided to several sub-applications as follows.

- Precision agriculture
    - Land-cover mapping and classification

    - Crop health monitoring
    - Modeling biophysical attributes
    - Studying soil characteristics
- Natural disaster management
    - Thermal disasters
    - Ground displacement
    - Floods
- Aquatic ecosystem management
    - Mapping and monitoring aquatic species
    - Characterizing water bodies
        - Channel bathymetry
        - River tracking
        - Characterizing thermal properties
- Polar remote sensing
- Wildlife research
    - Bird and mammal detection

First, the methodology and achievements of the studies are presented. Then, the general challenges, issues, and shortcomings as well as the future perspective of the applications are discussed. At the end, a general discussion is provided which gives an inclusive insight to the current characteristics and future requirements of unmanned aerial systems with respect to these applications.

Recent Applications of Unmanned Aerial Imagery in Natural Resource Management

by Mozhdeh Shahbazi, Jérôme Théau and Patrick Ménard

GIScience & Remote Sensing,

ISSN: 1548-1603

Volume 51, No. 4, Aug 2014, pp. 339-365

Abstract

Unmanned aerial vehicles have become popular platforms for remote sensing applications, particularly when spaceborne technology, manned airborne techniques, and *in situ* methods are not as efficient for various reasons. These reasons include the temporal and spatial data resolutions, accessibility over time and space, cost efficiency, and operational safety. Given that most commercial developers tend to focus on the hardware development of unmanned aerial systems, less attention is paid to the development and evaluation of their data processing techniques. Therefore, critical reviews of previous studies are required to describe the current state of research using data from unmanned remote sensing platforms. Accordingly, this paper presents the results of a comprehensive review of applications of unmanned aerial imagery for the management of agricultural and natural resources. This review attempts to demonstrate that developing robust methodologies and reliable assessments of results are significant issues for successful applications of unmanned aerial imagery.

Resumé

Les véhicules aériens sans pilote sont devenus des plateformes populaires pour les applications de télédétection, en particulier lorsque la technologie spatioportée, les techniques aéroportées habitées, et les méthodes terrestres ne sont pas assez efficaces pour diverses raisons. Ces raisons incluent notamment les résolutions et accessibilité temporelles et spatiales des données, les couts et la sécurité d'opération. Étant donné que la plupart des développeurs commerciaux ont tendance à se concentrer sur le développement matériel des systèmes aériens sans pilote, moins d'attention est accordée à l'élaboration et à l'évaluation des techniques de traitement des données. Par conséquent, un examen critique des études antérieures est nécessaire pour décrire l'état actuel de la recherche en utilisant des données acquises à l'aide des véhicules aériens sans pilote. Cet article présente donc les résultats d'une revue exhaustive des applications d'imagerie aérienne sans pilote pour la gestion des ressources naturelles et agricoles. Cette revue tente de démontrer que le développement de méthodes robustes et d'évaluations fiables des résultats constituent des aspects critiques pour garantir le succès des applications utilisant l'imagerie aérienne acquise par véhicule aérien sans pilote.

Keywords: unmanned aerial system; natural environment; agriculture; image processing; UAV;

2.2 Introduction

An unmanned aerial system (UAS) is a system comprising a number of sub-systems. These sub-systems include the unmanned aerial vehicle (UAV), vehicle launch (if applicable), control station (terrestrial, marine or aerial), and sensor payload (Gertler 2012).

As with many other remote sensing (RS) systems, the capabilities of UASs were initially explored by military organizations. They were used for different purposes, such as developing weapon platforms and acquiring data for tactical and strategic applications (Watts, Ambrosia, and Hinkley 2012; Austin 2010).

In most environmental RS applications, having frequent, quick and/or immediate access to updated data is mandatory. Moreover, the nature of some applications necessitates the use of unmanned control considering the safety, accessibility, and flexibility criteria (Watts, Ambrosia, and Hinkley 2012). UAVs have been identified as secure, lightweight, flexible, and automated platforms for such applications. Therefore, researchers have begun to show growing interest in deploying UASs for various civil applications including, but not limited to, meteorology, precision agriculture, wildlife research, forestry, land management, infrastructure inspection, traffic monitoring, epidemic emergencies, natural disaster management, and wilderness search and rescue.

Currently, there are many commercial developers who focus on the hardware development of unmanned aerial systems. However, less attention has been paid to the development and evaluation of data processing techniques. Therefore, critical reviews of previous studies are required to assess the current technological and scientific status of unmanned RS applications.

The unprecedented capabilities of small UAVs as aerial, low-altitude imaging platforms have not gone unnoticed by researchers. Digital image processing and interpretation as well as photogrammetric techniques can be applied to remotely acquired images, which lead to various sorts of geospatial information (Khorram et al. 2012). Examples of aerial imagery acquired for different purposes are presented in Figure 2.2.1.

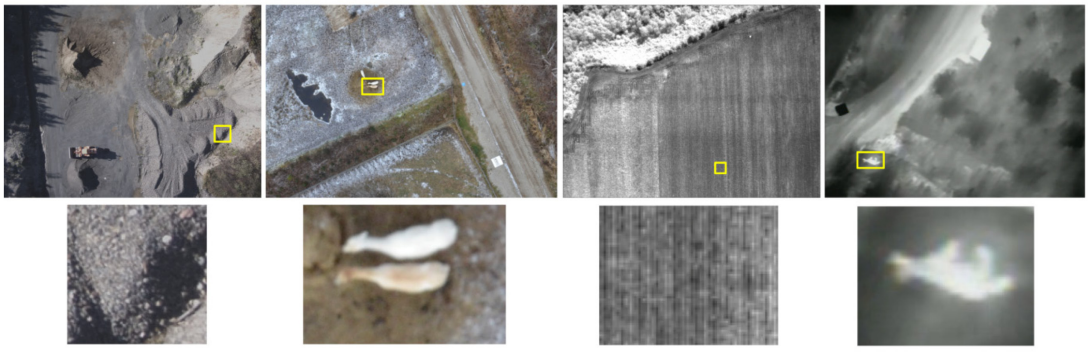


Figure 2.2.1. Examples of low-altitude, unmanned aerial imagery. From left to right: visible imagery for gravel-pit mapping, visible imagery for wildlife research, near-infrared imagery for precision agriculture, thermal-infrared imagery for wildlife research (Source: Université de Sherbrooke and Centre de géomatique du Québec).

In this regard, this article presents a comprehensive review of applications of unmanned aerial imagery for the management of natural resources. In particular, we focus on image-based processes; we also describe the present state of applications of unmanned aerial imagery, including the challenges and future opportunities. The articles analyzed in this review were selected based on the following criteria:

- The platform of the main data acquisition system comprised an unmanned aerial vehicle or a collaborative configuration of several UAVs.
- The end application was directly related to the management of agriculture and/or the natural environment.
- The payload consisted of a passive imaging sensor, and post-processing was applied to meet the requirements of the end-users.
- The study was not focused, merely, on gathering images, digital surface/terrain modeling and/or manned visual interpretation. Additionally, it proposed and evaluated a chain of automatic image-based processes to meet the application-related objectives.
- The study is considered as a scientific publication, such as an accessible[1] journal paper, a conference proceedings article or an official thesis.
- The study was performed in the 21st century, thus reflecting recent developments.

Figure 2.2.2 depicts the number of studies versus the publication date[2]. It indicates that there has been a considerable, growing interest in applying unmanned aerial imagery for managing agriculture and the natural environment in the last few years.

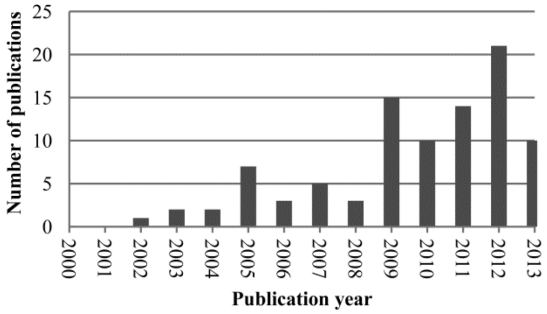


Figure 2.2.2. Number of reviewed articles versus publication date.

The applications of the reviewed articles can be categorized into five principal fields: precision agriculture and rangeland monitoring, natural disaster management, aquatic ecosystems management, polar remote sensing and wildlife research. Figure 2.2.3 compares the fields of the reviewed applications. Note that the statistics mentioned in this paper are calculated based on the reviewed articles. The reviewed applications are discussed in detail in the following sections (sections 2.3-2.7).

[1] Either open access or accessible through the libraries and archives of the Université de Sherbrooke

[2] The publications for 2013 are limited to those published by April.

Section 2.8 describes the overall characteristics of the reviewed systems; and, the conclusion is presented in section 2.9.

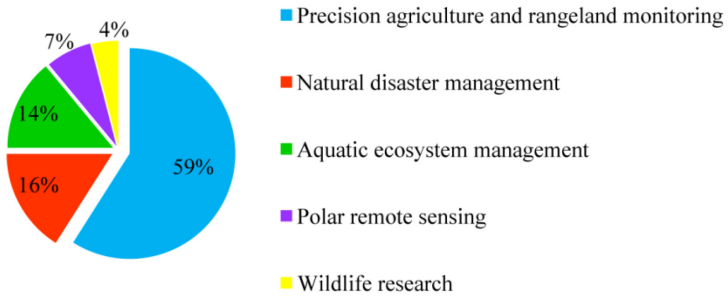


Figure 2.2.3. The applications of unmanned aerial imagery in the reviewed studies.

2.3 Precision agriculture and rangeland monitoring

Precision agriculture (PA) and rangeland monitoring have been the most common applications of unmanned aerial imagery (Zhang and Kovacs 2012). The following sections describe these applications and the image-based processes applied in the respective studies.

2.3.1 Land cover mapping and classification

Several studies have involved mapping and classifying land cover, based on various factors obtained from spectral data. Normalized difference vegetation index (NDVI) maps have frequently been used to generate field-scale green cover maps, to classify vegetated areas, and to segment the soil surface (McGwire et al. 2013; Primicerio et al. 2012; Arnold et al. 2010; Suzuki et al. 2010; Gay et al. 2009; Sugiura, Noguchi, and Ishii 2005). To be applicable in management practices, an NDVI map should be produced from geo-referenced and radiometrically adjusted images. Anisotropic reflection properties of the field can also be determined prior to flight and be applied for radiometric image correction (Grenzdorffer and Niemeyer 2011).

Various classification techniques have been applied in different studies to classify land cover. To classify agricultural surfaces where plants are spectrally similar, careful selection of classifiers is required. Feature descriptors based on texture and color were used by Reid, Ramos and Sukkarieh (2011) and Bryson et al. (2010) for classifying different types of vegetation on a farm. Object-oriented techniques of classification provide useful tools to deal with images with high spatial resolution for performing species-level classification (Laliberte and Rango 2011; Laliberte et al. 2011). Optimizing the segmentation scale and determining suitable features, based on texture, structure, shape, color, and other spectral properties, are the most significant challenges of these techniques (Laliberte and Rango 2009). In addition to the achievable accuracy for species-level vegetation classification (more than 85%), the transferability of classification rule sets is another advantage of object-oriented techniques (Laliberte et al. 2011). In other words, once a rule set is adjusted to certain types of species, it can be applied to similar sites with small changes in the training data and particular thresholds.

Other methods of land cover segmentation and unsupervised clustering have also been developed. In particular, these techniques were used for clustering tree species (Gini et al. 2012), mapping vigor differences (Johnson et al. 2003), segmenting catchment vegetation cover (Wundram and Loffler 2008) and measuring bare ground in rangelands (Breckenridge and Dakins 2011). Gini et al. (2012) concluded that the unsupervised clustering technique is efficient for differentiating segments with substantial spectral differences, such as roads and vegetation cover. However, it is not applicable to distinguish various tree species in the study area, and human interpretation is required to re-classify the results. Therefore, supervised classification is recommended instead. Wundram and Loffler (2008) observed a high level of miss-classification, even via supervised classification. The main reason was the similarity of the spectral signatures of the species in the visible range. In this case, integrating textural and topographic information can improve the results.

2.3.2 Crop health monitoring

Crop health monitoring is one of the most popular topics in precision agriculture via unmanned aerial imagery. This application can include disease detection, dead leaf detection, weed cover mapping, herbicide application monitoring, and water stress detection.

Various crop health descriptors should be combined with classification techniques to extract the diseased areas of a field. When the aerial system is only equipped with a sensor capturing visible light, textural features (such as angular moment, entropy, and contrast), fractal signatures, and color can be used to extract crop health factors (Samseemoung et al. 2012; Yue et al. 2012; Doudkin et al. 2009; Aber, Eberts, and Aber 2005). Otherwise, spectral information at other wavelengths can be applied as well. Xiang and Tian (2011a) presented a method to monitor a turf grass field where herbicide was applied, by temporally analyzing the changes in NDVI maps. They reported a slight difference (less than 2%) between herbicide damage areas estimated from imagery versus ground survey measurements. The analyses performed by Garcia-Ruiz et al. (2013) indicated that reflectance at 710 nm and NIR-R[1] index values were very different for healthy and HLB[2]-infected trees. They showed that the support vector machine (SVM) classification with non-linear kernel fitting had the highest accuracy for detecting diseased areas.

2.3.3 Modeling biophysical attributes

A considerable number of publications have been dedicated to modeling biophysical attributes of crops using unmanned aerial images as an alternative to destructive *in situ* measurements. The high spatial resolution of the images allows such parameters to be studied at various scales, for example at the farm scale, tree scale, and even finer scales such as the leaf scale.

The common way to model biophysical attributes of crops is to analyze the correlation between different vegetation indices and ground-based samples. Various techniques can be applied for this purpose, such as partial least squares estimation (Jensen et al. 2007), discriminant function analysis (Garcia-Ruiz et al. 2013), linear regression (Felderhof and Gillieson 2011), and support vector regression (Honkavaara et

[3] Near infrared and red

[2] Huanglongbing

al. 2012). It is worth noting that vegetation indices have temporal dependencies as well. Therefore, the season and the time of day when an index is derived and used for predicting a crop attribute should carefully be considered by analyzing temporal correlations (Xiang 2008). Table 2.3.1 provides a list of the attributes that were modeled in the reviewed articles.

Table 2.3.1. Summary of the studies on crop attribute modeling via unmanned aerial imagery.

Modeled attribute	**Airborne information**	**Crop**	**Result***	**Reference**
Productivity				
Yield Chlorophyll fluorescence	NDVI PRI[1]	Vineyard	N/A[2]	(Turner, Lucieer, and Watson 2011)
Grain yield, Protein level	DVI[3]	Wheat crop	R^2=0.91 R^2=0.66	(Jensen et al. 2007)
Yield and total biomass	NDVI	Rice crop	R^2=0.73 R^2=0.76	(Swain et al. 2010)
Dry biomass	NGRDI[4]	Soybean, alfalfa and corn crop	R^2=0.39, p=0.022 R^2=0.47, p=0.00034 R^2=0.88, p=0.00015	(Hunt et al. 2005)
Yield	Color	Orange orchard	R^2=0.3	(Schueller et al. 2006)
Canopy height	Color, topography	Buckwheat crop	R^2=0.87, p<0.001[5]	(Murakami et al. 2012)
Herbaceous biomass, Circlets phenomenon	Color	Semiarid farm	R^2=0.47, p<0.001	(Arnon et al. 2007)
Ripe coffee yield	Color	Coffee field	positive correlation (p<0.01)	(Herwitz et al. 2002)
Ripeness state of coffee cherries	Airborne reflectance at different wavelengths	Coffee field	Less than 18% difference between aerial prediction and parchment data	(Furfaro et al. 2005)
Rate of cross-pollination	Topography, Color	Maize crop	R=0.14 to 0.64, p<0.01	(Vogler and Eisenbeiss 2009)
Nutrients and LAI				
Total nitrogen and yield	Color, vegetation indices	Wheat crop	R^2>0.40	(Yunxia et al. 2005)
Leaf nitrogen level	$CCCI_{700,800}$[6]	Macadamia orchard	R^2=0.29, p=0.008	(Felderhof and Gillieson 2011)

[1] Photochemical reflectance index

[2] Not available

[3] Difference vegetation index

[4] Normalized green–red difference index

[5] Correlation of the measured and the estimated canopy height at sample points

[6] Canopy chlorophyll content index

Modeled attribute	Airborne information	Crop	Result*	Reference
Leaf nitrogen content	DGCI[1]	Rice crop	R^2>0.80	(Zhu et al. 2009)
Total nitrogen uptake per square meter,	GNDVI[2] ,	Wheat crop	R^2=0.92	(Lelong et al. 2008)
Leaf area index (LAI)	NDVI		R^2=0.82	
LAI	GNDVI	Wheat crop	R^2=0.85 (for LAI between 0 to 2.7)	(Hunt et al. 2010)
LAI	Canopy cover	Onion crop	R^2=0.75 to 0.84	(Corcoles et al. 2013)
Pigments				
FIPAR[3]	NDVI	Orchard	R^2>0.97	(Guillen-Climent et al. 2012)
Chlorophyll content	NDVI	Soybean crop	R^2>0.76	(Samseemoung et al. 2012)
Chlorophyll density	Red-Edge, NIR chlorophyll indices (CIs)	Rice crop	Strong correlation of field readings and estimated CIs (up to R^2=0.8)	(Uto et al. 2013)
Leaf carotenoid, Leaf chlorophyll content	R_{515}/R_{570}, TCARI/OSAVI	Vineyard	R^2=0.75 to 0.84, p<0.01 R^2=0.51, p<0.001	(Zarco-Tejada et al. 2013)
Water				
Stress status	Fluorescence emission	Orchard	R^2>0.54	(Zarco-Tejada et al. 2009)
Moisture content	NDVI	Wheat crop	R^2=0.37 to 0.78	(Han-Ya et al. 2010)
Canopy stomatal conductance, Leaf water potential, Water stress index,	Temperature	Orchard	R^2=0.78, p<0.05 R^2=0.34, p<0.001	(Zarco-Tejada et al. 2012; Berni et al. 2009)
Leaf stomatal conductance, Stem water potential (short-term response)	Temperature	Vineyard	R^2=0.68, p<0.01 R^2=0.50, p<0.05	(Baluja et al. 2012)
Leaf stomatal conductance, Stem water potential (long-term response)	TCARI /OSAVI[4] NDVI	Vineyard	R^2=0.84, p<0.05 R^2=0.68, p<0.05	(Baluja et al. 2012)
Water stress indicators (fruit quality parameters, titratable acidity and total	$PRI_{570,515}$	Citrus orchard	R^2=0.58 to 0.69	(Stagakis et al. 2012)

[1] Dark green color index

[2] Green normalized difference vegetation index

[3] Fraction of intercepted photosynthetically active radiation

[4] Ratio between transformed chlorophyll absorption in reflectance and optimized soil-adjusted vegetation index

Modeled attribute	Airborne information	Crop	Result*	Reference
soluble solids)				

*Unless otherwise stated, results are presented as the correlation coefficients (R^2) and their corresponding p-values between the crop attributes and the airborne information.

2.3.4 Studying soil characteristics

In addition to the vegetation, soil characteristics have also been studied from unmanned aerial images. Corbane et al. (2012) developed a two-scale classification technique to determine the soil surface characteristics (SSC) classes for a Mediterranean vineyard. The technique was based on multiple structural attributes of soil such as micro-topography, surface crusting, and soil cover. An average accuracy of 73% was achieved via mono-temporal classification, which was significantly improved to 82% via multi-temporal classification.

The soil erosion lines in agricultural areas were studied by D'Oleire-Oltmanns et al. (2012) via processing the digital terrain models (DTM) of a gully system. The DTMs were generated from unmanned aerial imagery and resulted in the horizontal and vertical accuracies of 0.9-2.7 and 1 cm, respectively. Although the DTMs were successfully applied to quantify the gully volume changes, an unavoidable minimum altitude of 250 meters restricted the spatial resolution of the imagery. In this case, this altitude was necessary to maximize the area coverage and minimize the error of camera orientation. The DTMs were processed to be applicable for hydrological modeling applications as well.

A similar study for gully measurement in badlands was performed by Gimenez et al. (2009). Although the error of photogrammetric measurements was as low as 3-12% for wide gullies, the accuracy was highly correlated with the width/depth ratio and morphology of the gullies (R^2=0.98). Depending on the time of day, the effect of sun-shadowing was the main reason for decreased accuracy when measuring deep, narrow gullies.

2.3.5 Challenges and future perspectives

Land cover classification and mapping applications are mainly concerned with typical sources of imaging errors, such as differential atmospheric attenuation of light rays between the ground and cameras, transmission differences caused by the wavelength response of the camera sensors, radiometric changes between overlapping images, shadows, camera vibrations, and forward motion artifacts. These errors must be considered by careful techniques of spectral calibration, atmospheric models, radiometric block adjustment, and motion compensation (Samseemoung et al. 2012; Primicerio et al. 2012; Grenzdorffer and Niemeyer 2011; Doudkin et al. 2009; Gay et al. 2009; Johnson et al. 2003).

Furthermore, the geometric distortions caused by inaccurate image registration and geo-referencing can be a considerable source of error. Due to such distortions, training samples cannot be precisely located on images. As a result, the extracted spectral information does not exactly correspond to the samples, and the estimates of biophysical parameters become erroneous (Honkavaara et al. 2012). In any case, it is believed that modeling crop attributes from unmanned aerial images yields more accurate results in

comparison with spaceborne and manned airborne imagery (Jensen et al. 2007). This is especially due to the accessibility to higher spatial resolution data (Zhu et al. 2009; Stagakis et al. 2012).

UAV-based PA systems are supposed to be more user-friendly and more autonomous, in part to attract farmers' interest (Zhang and Kovacs 2012). It is believed that more sophisticated techniques of classification, model estimation, change detection, and temporal analysis can be integrated into PA systems. These systems will be more autonomous if they don't require expert interpreters. The outcomes of experimental studies should also be evaluated for different case studies, for large areas, and against extensive *in situ* measurements to demonstrate their reliability to the end-users.

2.4 Natural Disaster Management

Natural disasters take various forms (e.g. floods, earthquakes, volcanic eruptions), and are one of the main concerns of governments and people all over the world. Remote sensing data, especially satellite data, are widely used to manage and monitor natural hazards at national and international scales (Tralli et al. 2005). This popularity is due to their wide coverage, spectral resolution, safety, and rate of update (Joyce et al. 2009; Gillespie et al. 2007). Unmanned aerial systems are also being considered to merge these RS systems for natural disaster management.

The studies that we reviewed can be categorized into two main groups of thermal disasters and ground displacement monitoring. These applications are described in the following sections.

2.4.1 Thermal disasters

Generating wildfire geo-located alarms and tracking fire parameters are the most common applications of unmanned aerial imagery for disaster management. Fusion of optical and thermal data is the principal way to provide information before, during and after wildfires. This information is mainly about anomalous heat levels (before an event) (Malos et al. 2013), combustion parameters (during the hazard), and hotspots and damage extent (after the event) (Ollero, Martínez-de-Dios, and Merino 2006).

The difference of reflectance values between fire pixels, burnt surfaces, and undamaged land cover is used as the key element for optical image processing at wildfires. Different techniques of thresholding and clustering are applied to individual spectral bands, spectral ratios, and extracted indices to identify hazard components (Merino et al. 2006). Thermal data are also fused with optical data for performing a more robust detection (Martinez et al. 2011; Pastor et al. 2011; Ambrosia and Wegener 2009). As mentioned by Ambrosia et al. (2003) and Esposito et al. (2007), careful calibration of thermal imaging scanners has a noticeable impact on increasing the accuracy of extracted information. Pre-processing for radiometric enhancement via noise cancelation and motion compensation is also suggested.

When the information is extracted from images, the detected segments should be geo-located. Using this information, the geometrical features of the event can be measured. These features include fire front location, fire base width and perimeter, flame length and height, inclination angle, coordinates of burnt areas, and location of hotspots (Martinez et al. 2011; Casbeer et al. 2006; Casbeer et al. 2005). Martinez et al. (2011) reported a mean relative error of 8% between the image-based locations of fire and the field

measurements. The main sources of error are related to the inaccuracy of 3D information and the smoke occluding the fire scene. The former mainly depends on the method of geo-referencing. The latter can be addressed by static cameras (both in the visible and infrared ranges) and complementary UASs.

Volcanic eruptions can be considered as another thermal disaster, which can be monitored by UASs. Visible and thermal images were used by Smith et al. (2009) to monitor an erupting volcano. The imagery was applied to derive information about changes in the active lava dome and vent, the collapse of the new lava dome, and the extent of hot areas.

Producing geo-coded mosaics of the hazard area is also important to crisis management systems. The extracted information for the thermal disaster can be demonstrated on the mosaics for visual analyses (Van Persie et al. 2011; Zhou 2009). The most common method to generate the mosaics is to geo-reference the images individually and then mosaic them all together (Xiang and Tian 2011b).

There are several methods for geo-referencing unmanned aerial images. Direct geo-referencing via the navigation data is the most time-effective way to automatically carry out the task. Xiang and Tian (2011b) reported an average accuracy of 0.56 m for the mosaics produced with 5 cm spatial resolution. In addition to the accuracy of navigation sensors, calibrating the platform and internal camera orientation are important factors to improve the accuracy (Mostafa and Schwarz 2001; Habib et al. 2006).

In wildfires, the density and height of trees can affect the performance of global positioning systems (Dussault et al. 1999). Therefore, indirect geo-referencing via absolute space resection can be performed as an alternative (Zhou 2009). Planimetric maps, topographic maps, digital elevation models (DEMs), and ground control points (GCPs) for the study area can be used for this purpose.

Turner, Lucieer, and Watson (2012) compared the direct and indirect methods of geo-referencing the point clouds. Using those point clouds, individual images were rectified and stitched to form the mosaics. They showed that mosaics with average accuracies of 65-120 cm and 10-15 cm can be generated from the direct and indirect methods, respectively. The accuracy of indirectly created mosaics depends on the accuracy of relative modeling (point cloud generation), positioning the GCPs, and registering them to the images.

2.4.2 Ground displacement

UAV technology provides an approach for a quick response to disastrous phenomena caused by ground displacements, such as landslides and earthquakes. Several sorts of data can be used to detect ground displacements and changes associated with them. The most useful type of data is the topographic data in the form of DEMs and ortho-photos. High-resolution images can be efficiently used for both topographic data generation and change detection (Niethammer et al. 2012).

In this regard, Gong et al. (2012) developed an object-oriented classification technique to derive river channel information, and to perform change analysis on images taken before, during and after an earthquake. The changes in the pattern and area of the channels, the river width, and the width of the landslide orifice perpendicular to the river were analyzed by comparing the results of the classifications. The runways were extracted with a maximum error of four pixels before the disaster and eight pixels

after the disaster. Therefore, the classification could be accurate enough to provide the basis for change detection purposes.

To monitor landslides, it is necessary to perform spatial observations, temporal tracking of surface changes, and the timely distribution of the information to crisis management units (Van Persie et al. 2011). Applying automated processing techniques to UAV imagery can reduce the level of manual interpretation and increase the time efficiency for widespread landslide mapping.

Color information, slope gradients, and vegetation indices are the features that can be derived from high-resolution optical images (Rau et al. 2011); these features are used with classification techniques to differentiate landslides from other areas. The contour-line behavior of landslides is another key factor that helps their detection from topographic maps (Tahar, Ahmad, and Akib 2011). In their simulated experiments, they were able to calculate the volume of soil loss caused by landslides with 2% error. Optical images may also be fused with other data sources, such as deformation maps from LiDAR[1] or DInSAR[2], to map the extent of damage after an event (Kim et al. 2012).

Slope stability analysis at dams, embankments, mines, and natural slopes is an important issue for land exploitation planning. The main consequence of slope instability is terrain displacement in the form of landslides, landslips, and ground subsidence. Therefore, monitoring the stability of slopes at risky places is an essential task to prevent such disasters.

Photogrammetric techniques can be applied to unmanned aerial images to generate accurate slope maps (Tahar et al. 2012). These maps can be analyzed to acquire more sophisticated information about the slope characteristics. An example of stability analysis was performed by Firpo et al. (2011) using distinct element numerical methods. They derived precise information about discontinuities and slope geometry from optical images, even in high fronts that were inaccessible for direct measurements. They concluded that the results were very accurate due to the quality of stereo-mapping.

Detection of sub-decimeter surface fissures and faults is important as an indicator of hazards. In this regard, Stumpf et al. (2012) proposed a technique to detect surface cracks from aerial images. Their image-processing algorithm was based on Gaussian matched filters and first-order derivates in a multi-scale framework. Automatic techniques of fissure detection eliminate the need for exhaustive expert work to locate the fissures with variable sizes at different scales. There are different techniques of edge-detection and linear feature extraction that can be applied in this context. The examples include derivative edge detection techniques (Roberts, Prewitt, Sobel, Laplacian, Laplacian of Gaussian, canny, and moment-based operators), and non-derivative techniques based on principal component analysis and interpolation (Bhardwaj and Mittal 2012; Lin, Jiang and Wang 2011; Hermosilla et al. 2008). Since edges are sensitive to image noise, techniques of noise removal and image enhancement should be considered in the pre-processing step (Bhardwaj and Mittal 2012). Post-processing is also required to distinguish regular edges from those that indicate fissures or cracks.

[1] Light detection and ranging

[2] Differential synthetic aperture radar interferometry

2.4.3 Floods

Flooding is another event that can be monitored remotely by integrating the techniques of water body detection, vegetation mapping, DEM generation, and hydrological modeling. Satellites sensors in the visible, infrared, and microwave range are currently used to monitor rivers and to delineate flood zones (Syvitski et al. 2012; Yilmaz et al. 2010).

However, flood mapping is a considerable challenge for spaceborne passive imagery. This is mostly due to the presence of dense cloud cover, closed vegetation canopies, and the satellite revisit time and viewing angle (Joyce et al. 2009; Snyal and Lu 2004). Applying UAVs with an appropriate flight mode provides the possibility to overcome these issues. Therefore, monitoring inundation and measuring hydrological parameters of floods can be considered as future applications for UASs.

2.4.4 Advantages and issues of UASs

Thermal disasters (wildfires being of most concern) are very dynamic phenomena. They spread with varying rates that are affected by different environmental factors, such as soil moisture, vegetation type, and weather conditions (Bowman et al. 2009). Considering these characteristics, frequent and updated information about fire parameters is essential to assist management teams in arranging fire fighting plans.

Although satellite imagery provides a frequent, global overview of wildfires, fine-scale tactical management operations still require higher spatial resolution (Ambrosia and Wegener 2009). In addition, manned airplanes are not capable of flying safely in the presence of dense smoke, for long stretches of time through the night and day, and without changing the crews (Pastor et al. 2011). Comparatively, unmanned aerial systems can be configured for safe flights for both surveillance and fire fighting schemes.

Considering the extension of natural fires, a cooperative configuration of UASs and terrestrial sensors is very useful to gather all the necessary details of wildfires (Martinez et al. 2011). This indicates the necessity of designing very sophisticated management systems, sensor integration techniques and data fusion methods. It is also believed that autonomous control of UAVs is less recommended for disaster management applications, in comparison with an experienced remote pilot. The main reason for this is that autonomous flight mode is less able to deal with uncertain conditions such as gusty winds or smoke haze (Niethammer et al. 2012).

For disaster management, accurate and automatic generation of topographic products, such as DEMs and ortho-rectified mosaics, is of great significance (Li et al. 2011; Chou et al. 2010; Lin et al. 2010; Wu and Zhou 2006; Qian et al. 2012). However, it requires the development of image processing chains, which would reduce the influence of analysts' interpretations and time-consuming man-computer interactions.

When UAS-based mapping is used for ground displacement measurements, considering the following factors can significantly improve the accuracy (Honkavaara et al. 2013; Honkavaara et al. 2012; Hruska et al. 2012; Eisenbeiss and Sauerbier 2011; Hernández López et al. 2011; Lelong et al. 2008).

- The mapping techniques should take into account the defects of consumer-grade, optical elements and sensors integrated into small-scale UAVs (both radiometric and geometric).
- They must produce geo-referenced mosaics, which can replace the enormous number of images produced during each flight due to the small format of imaging sensors.
- They should consider the specifications of low-altitude imagery, such as large geometric distortions, overlap inconsistency, variable motion blur, low contrast of the objects, non-uniformity of the distribution of features, large rotation and scale variations among images, high terrain relief relative to flight altitude, and non-uniform block illumination.

2.5 Aquatic Ecosystem Management

2.5.1 Why UAVs?

Remote sensing systems are fast, non-intrusive alternatives to ground-based approaches for mapping and monitoring aquatic ecosystems such as wetlands, swamps, bogs, riparian forests, and rainforests. However, the task may involve identifying vegetation types in detail and distinguishing small plant species, which require high spatial resolution, even of less than one centimeter (Ishihama, Watabe, and Oguma 2012). Low-altitude UASs can satisfy this requirement. Moreover, their ability to fly below cloud cover and their portability make them suitable for all seasons of the year (Ishihama, Watabe, and Oguma 2012; Fletcher and Erskine 2012; Koh and Wich 2012; Li et al. 2010).

The following two sections discuss the use of UAV imagery for aquatic ecosystem management in regards to two main aspects: mapping aquatic species and characterizing water bodies.

2.5.2 Mapping and monitoring aquatic species

Detecting specific species is important for conservation decisions in order to monitor particular communities, assemblages, or habitats. Various classification techniques have been applied to unmanned aerial images to identify particular structures in aquatic ecosystems. Table 2.5.1 presents a summary of these studies with respect to the general methods applied and their main objectives.

The main challenge of these applications is that, for most aquatic zones, lack of brightness and low contrast in the vegetation cover make land cover classification a difficult task. Therefore, imaging at wavelengths other than visible range, such as infrared and near infrared, should be used to improve the contrast of vegetation classes. Moreover, as the variance in pixel values increases with the resolution, identifying the right class for every pixel becomes more difficult than classifying the image segments (objects) (Lechner et al. 2012). Consequently, object-oriented classification techniques are preferred (Strecha et al. 2012; Knoth et al. 2013). Dunford et al. (2009) compared pixel-based and object-oriented classification techniques to classify riparian vegetation units. They evaluated both approaches for single images and for mosaics. As their results indicate, object-oriented classification provides higher accuracy than pixel-based classification for single images. Comparatively, classifying at the mosaic level is simpler, since mosaics cover a large area and accelerate the surveying process; besides, fusing classification results obtained from single images is not required. However, for mosaics, the classification accuracy greatly depends on the variation of the illumination over the image series and on

the geometric distortions. Therefore, object-oriented classification of mosaics is only suggested when radiometric inconsistency of images is adjusted, and geometric errors are duly rectified.

Table 2.5.1.Summary of the applications of unmanned aerial imagery in mapping and monitoring aquatic species

Objective	General method	Result	Reference
Detecting a specific type of swamp fern (Gleichenia)	Maximum likelihood classification	N/A	(Strecha et al. 2012)
Classifying the cutover bog surface structures	Object-oriented classification based on NIR reflectance, texture, shape	Overall classification accuracy = 91%	(Knoth et al. 2013)
Detecting the boundary between the swamp and its surrounding eucalypt woodlands	Object-based classification based on surface height and image edges	Classification accuracy = 95-100%	(Lechner et al. 2012)
Identifying the expansion of wetland invasive grass species	Classification based on multiclass relevance vector machines	Classification accuracy = 95%	(Zaman et al. 2011)
Detecting aquatic weed infested areas	Supervised classification based on support vector machines	N/A	(Goktogan et al. 2010)
Mapping macro-algal cover, Modeling macro-algal dry weight	Supervised classification, Linear regression with: NDVI, NGRDVI, Saturation	N/A R^2=0.75, p=3.6e^{-8} R^2=0.68 to 0.83, p<1.7e^{-8} R^2=0.68 to 0.75, p<1.6e^{-7}	(Dugdale 2007)
Classify riparian vegetation units, Identifying dead wood and canopy mortality	Pixel-based classification base don decision-trees, Object-oriented classification based on color, texture, size, shape and topological relationships	Overall classification accuracy: for vegetation units= 63-91% for dead woods= 80% (regarding omission error) and 65% (commission)	(Dunford et al. 2009)

2.5.3 Characterizing water bodies

Spectral observations from UAVs are applied in various applications such as channel bathymetry, river tracking, and thermal characterization of aquatic ecosystems. A summary of these studies is presented in Table 2.5.2.

Table 2.5.2. Summary of the applications of unmanned aerial imagery for characterizing water bodies

Objective	General method	Reference
Channel bathymetry	Correlating the observed water depth samples with pixel gray values from RGB channels	(Lejot et al. 2007)
River tracking	Histogram-based NIR thresholding	(Han 2009)
Stream extraction	NIR and IR thresholding	(Jensen et al. 2012)

Water body extraction Channel classification Modeling water temperature distribution Analyzing discharge rates of the river and thermal variability patterns	Masking and neighborhood analysis Object-oriented fuzzy classification Gaussian probability density functions Temporal and spatial analysis of temperature models	(Wawrzyniak et al. 2013)

2.5.4 Channel bathymetry

To characterize river morphology and monitor river restoration, the bathymetry of the aquatic areas is required. Channel bathymetry from spectral data is based on the relationship among light waves, the water surface and the channel bed, which can be reduced to the radiometric signal only, in the case of homogeneous atmospheric, substrate and water column conditions (Legleiter et al. 2004). Lejot et al. (2007) concluded that water depths calculated from unmanned optical data (specifically red and blue bands) were highly correlated with field observations (R^2>0.53).

Several sources of errors such as geo-referencing, atmospheric conditions, and illumination conditions may affect the results of bathymetry from optical data. Median filtering, histogram matching and sub-grouping the images were used by Lejot et al. (2007) to eliminate these errors. However, these issues were not completely addressed and remain as considerable challenges of unmanned image-based bathymetry.

2.5.5 River tracking

River tracking from optical images is a simple way of acquiring a valuable source of information that can be applied to water resource management (Han 2009). Besides, the results of real-time river tracking can be used for vision-aided navigation purposes. However, rivers are texture-less structures with irregular boundaries and branches (Rathinam et al. 2007). Therefore, proper learning algorithms and integration with NIR sensors are required.

2.5.6 Characterizing thermal properties

Temperature is a significant variable of aquatic systems, as it has a close relationship with several environmental processes, such as biological reactions, chemical alterations and the spread of aquatic species (Beveridge, Petchey, and Humphries 2010). Thermal imaging from UAVs provides the capacity to characterize the thermal properties of aquatic ecosystems (Wawrzyniak et al. 2013). However, it is necessary to continuously correct the thermal images using terrestrial samples and atmospheric models (Wawrzyniak et al. 2013; Jensen et al. 2012).

2.6 Polar Remote Sensing

2.6.1 Advancements of UASs

During the last decades, noticeable progress has been made in RS techniques for observing and monitoring the polar regions (Jezek and Onstott 1999; Lubin and Massom 2007). Optical, thermal, microwave, and radar data gathered by low-orbit satellites are mainly used to monitor the environment of the polar regions (Lubin and Massom 2007; Stine et al. 2010; Lin 2008). Comparatively, aerial RS

systems and *in situ* techniques are not frequently applied. This is mainly due to the harsh environment of the polar regions. The areal coverage provided by spaceborne sensors is also very important in global studies of the polar regions.

Nevertheless, satellite data should mostly be fused with other data for both evaluation and enhancement purposes (Gruen 2000). Furthermore, aerial RS systems can be used to observe local environmental changes with higher spatial and temporal resolutions. For example, mapping ice cover, permafrost patterns, periglacial processes, and vegetation is possible via aerial remote sensing (Boike and Yoshikawa 2003). Unmanned aerial systems can facilitate these applications, particularly when using manned platforms can be expensive, risky, and cumbersome.

UAVs can be designed to endure the tough climate conditions of these regions and gather spectral data below the cloud cover. For instance, the ability of UAVs to fly accurately in waypoint navigation mode is useful for gathering complex data for measuring the bidirectional reflectance factor (BRF) of snow cover (Hakala et al. 2010). BRF retrieval is useful for both measuring surface albedo and enhancing satellite data. The data acquired from UAVs can also be applied to extract other pertinent information for regional environmental management. Such applications can benefit from either lightweight multi-spectral sensors or synthetic aperture radars (Koo et al. 2012).

2.6.2 Applications of unmanned aerial imagery over the polar regions

Measuring the volume of supraglacial lakes is a useful means of verifying the water balance of ice sheets. Satellite data are effective for observing supraglacial lakes. However, cloud cover, satellite overpass times and spatial resolution are the limiting factors, since the lakes have variable sizes, and they form and drain very quickly. Therefore, a combination of satellite and unmanned aerial images was suggested by Lettang et al. (2013). They used a supervised classification, based on spectral signatures from satellite data and textural features from unmanned imagery. Applying high-resolution aerial images was determined as the only way to analyze the texture of former lake areas and the effects of liquid water on ice texture.

The utility of unmanned aerial imagery obtained over the Arctic Ocean has been explored in the literature to characterize the variability of sea ice with melt ponds (Inoue, Curry, and Maslanik 2008; Curry et al. 2004). In comparison with unmanned aerial imagery, they showed that the resolution and accuracy of passive satellite imagery is too low to differentiate open water from ponded ice and to calculate ice concentration, especially during the melt season.

Lucieer, Robinson, and Turner (2010) developed a UAV-based photogrammetric system to extract the micro-topography of moss beds in order to study the regional effects of climate change. Thermal imagery was also acquired to map the heat regulation of bryophytes. The generated DEM was used to calculate the topographic wetness index as an indicator of the spatial distribution of potential surface wetness caused by snowmelt. Although a correlation between water availability and the extent of moss beds was observed, no analysis was performed to quantify this relationship.

2.7 Wildlife Research

Natural ecosystem management and conservation requires adequate monitoring of biodiversity as well as regular surveying of wildlife and species' status (Mackenzie 2005). Applications of UASs to detect birds, large mammals, and marine mammals are described in the following sections.

2.7.1 Bird and mammal detection

Unmanned remote sensing systems have recently been developed for wildlife monitoring purposes by Israel (2011) for detecting deer fawns, by Schoonmaker et al. (2008) and Koski et al. (2009) for tracking marine mammals and by Vermeulen et al. (2013) for counting elephants. In the studies mentioned, however, detection and counting procedures were performed by visual interpretation of real-time or offline images. The applications of UAV imagery for automatic detection of birds and marine mammals are summarized in Table 2.7.1.

Table 2.7.1. Applications of unmanned aerial imagery for wildlife research

Objective	Method	Reference
Detecting bird patterns	Template matching by normalized cross correlation and region-growing segmentation	(Abd-Elrahman 2005)
Counting the snow geese	Thresholding based on intensity and shape	(Chabot and Bird 2012)
Black-headed gull recognition Colony size estimation	Detection based on shape and color Active nest identification from sequential images Identifying nest structures from single images	(Sarda-palomera et al. 2012)
Detecting marine mammals	Morphological operations and color thresholding	(Hodgson et al. 2010)

2.7.2 Pros and cons of UASs for wildlife research

During recent decades, wildlife managers have applied aerial surveying as a tool for detecting, counting, tracking and monitoring animals. However, aerial surveys are associated with some concerns. First, crashes of low-altitude small aircraft have been the primary cause of work-related death for wildlife biologists. Secondly, the management procedures require frequent observations through time and space, which entail the huge cost of hiring planes and pilots. Thirdly, most wildlife spectral data contain statistical errors; this is principally due to the temporal and spatial resolution of the acquired data (Jones 2003).

In contrast, UAVs are proven to be safe and cost-effective with the possibility of being equipped with high-resolution imaging systems. Besides, due to their small size and low noise level, UAVs are capable of flying over flocks without disturbing them or causing them to escape (Vermeulen et al. 2013; Chabot and Bird 2012; Sarda-palomera et al. 2012). Additionally, UAVs have the capacity to follow track lines more accurately than manned aircraft (Hodgson et al. 2010); this is a significant characteristic for tracking animals.

Although UAVs offer enhanced opportunities as wildlife monitoring tools, on-going research is still required to derive useful information from the remotely gathered data (Jones, Pearlstine and Percival 2006). There are several issues, which should be considered and addressed to develop image-based processes for animal detection. The main issues include the spectral similarity of animals or birds to

their surrounding environment (Abd-Elrahman 2005), varying size of animals (adults and calves) (Vermeulen et al. 2013), and their displacement during the image series (Sarda-palomera et al. 2012). Therefore, the appropriate use of multi-spectral (especially thermal infrared) imaging, multi-scale classification techniques, motion estimation methods and target tracking algorithms should be considered to avoid these problems.

2.8 System Characteristics

A summary of the aerial systems deployed in the reviewed studies presents the general characteristics of their platforms and sensors (Table 2.8.1).

Table 2.8.1. The percentages of the reviewed systems with specific characteristics in each field of application

		Field of application					
		All	**PA and rangeland**	**Natural disaster**	**Aquatic ecosystem**	**Polar**	**Wildlife**
UAV Class	Fixed-wing	49.4	47.2	38.5	53.8	50.0	100.0
	Rotary-wing	39.3	41.5	53.8	30.8	33.3	0.0
	Other*	11.3	11.3	7.7	15.4	16.7	0.0
UAV engine power	Electric	48.1	36.2	66.7	71.4	33.3	100.0
	Fuel	35.4	44.7	22.2	14.3	50.0	0.0
	Solar	3.8	6.4	0.0	0.0	0.0	0.0
	Unpowered	12.7	12.7	11.1	14.3	16.7	0.0
UAV payload capacity	< 1 kg	33.3	37.2	33.3	25.0	16.7	50.0
	>1 & <5 kg	38.9	23.3	55.6	75.0	50.0	50.0
	> 5 kg	27.8	39.5	11.1	0.0	33.3	0.0
Camera spectral bands	Red	81.5	75.9	86.7	92.3	83.3	100.0
	Green	82.6	77.8	80.0	100.0	83.3	100.0
	Blue	77.2	66.7	86.7	100.0	83.3	100.0
	Near IR	31.5	35.2	6.7	53.8	33.3	0.0
	Thermal IR	21.7	14.8	53.3	15.4	33.3	0.0
	Multi-spectral**	9.8	16.7	0.0	0.0	0.0	0.0
	Hyper-spectral	7.6	9.3	6.7	0.0	16.7	0.0

* Airship, kite, balloon, glider

** Multiple spectral bands other than Red, Green, Blue, NIR and Thermal IR

Through this analysis, it was inferred that fixed-wing aircraft are more frequently used than rotary ones, and they are mostly powered by electric motors. Furthermore, most of the systems (more than 72%) are restricted by payload capacities lower than five kilograms. These characteristics can be considered either as an advantageous feature or as a drawback of these systems. The small and lightweight systems are, mostly, low-cost regarding both the vehicle price and the operation costs; besides, in the event of a crash, there will be less damage to property, people, and study fields. However, the payload capacity restriction affects the selection of the sensors; that is, less sophisticated sensor combinations can be

deployed. For instance, consumer-grade cameras are utilized instead of conventional aerial cameras. Less than 30% of the systems are capable of acquiring multi-spectral bands other than RGB. This could mainly be due to the weight restrictions that do not allow adding heavier sensors and the required energy sources. In the future, it is hoped that more lightweight, multi-spectral image acquisition systems with a lower cost of production can be designed for integration into UAV-based remote sensing systems. For example, the experiments by Gay et al. (2009), Jensen et al. (2007) and Hunt et al. (2010) indicate that, with careful treatment of filters and spectral calibration, commercial digital cameras can be modified to acquire other ranges of wavelengths in the electromagnetic spectrum. Readers are referred to Hardin and Hardin (2010) and Hardin and Jensen (2011) for further details and discussions on the characteristics of small-scale UAVs.

The quality and accuracy of measurements and observations with non-metric cameras can be greatly increased by geometric and spectral calibrations followed by radiometric corrections of the acquired images. However, more than 60% of the studies have reported no effort for spectral calibration and/or radiometric correction of images; besides, only 27% of them have geometrically calibrated the cameras.

Regarding the navigation sensors, only a few systems have been equipped with geodetic-grade navigation sensors. Most systems deploy a simple form of INS[1] assembled with a GPS. Differential GPS[2] measurement in the form of code-phase differential GPS or carrier-phase differential GPS, known as real time kinematic (RTK), improves the direct geo-referencing accuracy up to sub-meter and sub-centimeter scales. However, this requires more expensive and sophisticated equipment (Valavanis 2007). There are also other means of navigation, especially for GPS-denied environments. They include radar tracking, radio tracking, dead reckoning and vision-based navigation. The accuracy of mapping with these systems is limited by the accuracy of navigation data unless ground control observations are applied for indirect geo-referencing.

2.9 Conclusion

This review describes the current state of research on applications of unmanned aerial imagery to the management of agriculture and the natural environment. The main advantages, limitations, and future perspectives of these applications are also discussed.

The current applications are still in the initial stages of development. This could be due to the fact that UAV-based environmental remote sensing has not yet developed multi-disciplinary collaborations. Another reason is that current operations of UASs are limited due to regulatory issues (Rango and Laliberte 2010). Unmanned flight certifications are required for all research purposes to prevent the damage that might be caused by UAVs, for example, by crashing into people, colliding with other aircraft or by disrupting civil and military services.

[1] Inertial navigation system

[2] Global positioning system

Nevertheless, interest is rapidly growing in the applications of high-resolution, multi-spectral images acquired via UAVs. We expect that the expanding research and development of these systems will increasingly demonstrate their relevance to the management of agricultural and natural environments.

Acknowledgments

This study is supported in part by grants from the: Centre de Géomatique du Québec (www.cgq.qc.ca), Fonds de Recherche Québécois sur la Nature et les Technologies (www.frqnt.gouv.qc.ca), and Natural Sciences and Engineering Research Council of Canada (www.nserc-crsng.gc.ca). The authors would like to thank Dr. Catherine Brown for her assistance in language editing.

2.10 References

Abd-Elrahman, A. 2005. "Development of Pattern Recognition Algorithm for Automatic Bird Detection from Unmanned Aerial Vehicle Imagery." *Surveying and Land Information Science* 65 (1): 37-46.

Aber, J. S., D. Eberts, and S. W. Aber. 2005. "Applications of Kite Aerial Photography: Biocontrol of Salt Cedar (Tamarix) in the Western United States." *Transactions of the Kansas Academy of Science* 108 (1-2): 63-66. doi: 10.1660/0022-8443.

Ambrosia, V. G., and S. S. Wegener. 2009. "Unmanned Airborne Platforms for Disaster Remote Sensing Support." In *Geoscience and Remote Sensing*, edited by Pei-Gee Peter Ho, 91–114. Croatia: InTech Open Science| Open Minds.

Ambrosia, V. G., S. S. Wegener, D. V. Sullivan, S. W. Buechel, S. E. Dunagan, J. A. Brass, and S. M. Schoenung. 2003. "Demonstrating UAV-Acquired Real-Time Thermal Data over Fires." *Photogrammetric Engineering & Remote Sensing* 69 (4): 391-402.

Arnold, T., M. De Biasio, A. Fritz, A. Frank, and R. Leitner. 2010. "UAV-based Multi-spectral Environmental Monitoring." In *Proceedings of IEEE Sensors Conference*, 995-998. Kona, HI: IEEE.

Arnon, A. I., E. D. Ungar, T. Svoray, M. Shachak, J. Blankman, and A. Perevolotsky. 2007. "The Application of Remote Sensing to Study Shrub-Herbaceous Relations at a High Spatial Resolution." *Israel Journal of Plant Sciences* 55 (1): 73-82. doi: 10.1560/IJPS.55.1.73.

Austin, R. 2010. *Unmanned Aircraft Systems: UAVs Design, Development and Deployment*. Chichester: John Wiley & Sons.

Baluja, J., M. P. Diago, P. Balda, R. Zorer, F. Meggio, F. Morales, and J. Tardaguila. 2012. "Assessment of Vineyard Water Status Variability by Thermal and Multispectral Imagery Using an Unmanned Aerial Vehicle (UAV)." *Irrigation Science* 30 (6): 511-522. doi: 10.1007/s00271-012-0382-9.

Berni, J., P. J. Zarco-Tejada, G. Sepulcre-Canto, E. Fereres, and F. Villalobos. 2009. "Mapping Canopy Conductance and CWSI in Olive Orchards Using High Resolution Thermal Remote Sensing Imagery." *Remote Sensing of Environment* 113 (11): 2380-2388. doi: 10.1016/j.rse.2009.06.018.

Beveridge, O. S., O. L. Petchey, and S. Humphries. 2010. "Direct and Indirect Effects of Temperature on the Population Dynamics and Ecosystem Functioning of Aquatic Microbial Ecosystems." *Journal of Animal Ecology* 79 (6): 1324-1331. doi: 10.1111/j.1365-2656.2010.01741.x.

Bhardwaj, S., and A. Mittal. 2012. “A Survey on Various Edge Detector Techniques.” *Procedia Technology* 4: 220-226. doi: 10.1016/j.protcy.2012.05.033.

Boike, J., and K. Yoshikawa. 2003. “Mapping of Periglacial Geomorphology Using Kite/Balloon Aerial Photography.” *Permafrost and Periglacial Processes* 14 (1): 81-85. doi: 10.1002/ppp.437.

Bowman, D. M., J. K. Balch, P. Artaxo, W. J. Bond, J. M. Carlson, M. A. Cochrane, C. M. D’Antonio et al. 2009. “Fire in the Earth System.” *Science* 324 (5926): 481-484. doi: 10.1126/science.1163886.

Breckenridge, R. P., and M. E. Dakins. 2010. “Evaluation of Bare Ground on Rangelands Using Unmanned Aerial Vehicles: a Case Study.” *GIScience & Remote Sensing* 48 (1): 74-85. doi: 10.2747/1548-1603.48.1.74.

Bryson, M., A. Reid, F. Ramos, and S. Sukkarieh. 2010. “Airborne Vision-Based Mapping and Classification of Large Farmland Environments.” *Journal of Field Robotics* 27 (5): 632-655. doi: 10.1002/rob.20343.

Casbeer, D. W., R. W. Beard, T. W. McLain, S. M. Li, and R. K. Mehra. 2005. “Forest Fire Monitoring with Multiple Small UAVs.” In *Proceedings of IEEE American Control Conference*, 3530 - 3535. Portland: IEEE.

Casbeer, D., D. Kingston, R. Beard, and T. McLain. 2006. “Cooperative Forest Fire Surveillance Using a Team of Small Unmanned Air Vehicles.” *International Journal of Systems Science* 37 (6): 351-360. doi: 10.1080/00207720500438480.

Chabot, D., and D. M. Bird. 2012. “Evaluation of an Off-The-Shelf Unmanned Aircraft System for Surveying Flocks of Geese.” *Waterbirds* 35 (1): 170-174. doi: 10.1675/063.035.0119.

Chou, T. Y., M. L.Yeh, Y. C. Chen, and Y. H. Chen. 2010. “Disaster Monitoring and Management by the Unmanned Aerial Vehicle Technology.” In *Proceedings of ISPRS TC VII Symposium*, 137-142. Vienna: ISPRS.

Corbane, C., F. Jacob, D. Raclot, J. Albergel, and P. Andrieux. 2012. “Multitemporal Analysis of Hydrological Soil Surface Characteristics Using Aerial Photos: a Case Study on a Mediterranean Vineyard.” *International Journal of Applied Earth Observation and Geoinformation* 18: 356-367. doi: 10.1016/j.jag.2012.03.009.

Corcoles, J. I., J. F. Ortega, D. Hernandez, and M. A. Moreno. 2013. “Use of Digital Photography from Unmanned Aerial Vehicles for Estimation of Leaf Area Index in Onion (Allium Cepa L.).” *European Journal of Agronomy* 45 (Feb.): 96-104. doi: 10.1016/j.eja.2012.11.001.

Curry, J. A., J. Maslanik, G. Holland, J. Pinto, G. Tyrrell, and J. Inoue. 2004. “Applications of Aerosondes in the Arctic.” *Bulletin of American Meteorological Society* 85 (12): 1855-1861. doi: 10.1175/BAMS-85-12-1855.

Doudkin, A., V. Ganchenko, A. Petrovsky, and B. Sobkowiak. 2009. “Disease Areas Detection on Agricultural Plants Using Fractal and Textural Features of High Resolution Color Aerial Photographs.” *Journal of Research and Application in Agriculture Engineering* 54 (3): 55-59.

Dugdale, S. J. 2007. “An Evaluation of Imagery from an Unmanned Aerial Vehicle (UAV) for the Mapping of Intertidal Macroalgae on Seal Sands, Tees Estuary, UK.” Master of Science Thesis, Durham University.

Dunford, R., K. Michel, M. Gagnage, H. Piegay, and M. L. Tremelo. 2009. "Potential and Constraints of Unmanned Aerial Vehicle Technology for the Characterization of Mediterranean Riparian Forest." International Journal of Remote Sensing 30 (19): 4915-4935. doi: 10.1080/01431160903023025.

Dussault, C., R. Courtois, J. P. Ouellet, and J. Huot. 1999. "Evaluation of GPS telemetry collar performance for habitat studies in the boreal forest." *Wildlife Society Bulletin*, 27 (4): 965-972.

Eisenbeiss, H., and M. Sauerbier. 2011. "Investigation of UAV Systems and Flight Modes for Photogrammetric Applications." *The Photogrammetric Record* 26 (136): 400-421. doi: 10.1111/j.1477-9730.2011.00657.x.

Esposito, F., G. Rufino, A. Moccia, P. Donnarumma, M. Esposito, and V. Magliulo. 2007. "An Integrated Electro-Optical Payload System for Forest Fires Monitoring from Airborne Platform." In *Proceedings of IEEE Aerospace Conference*, 1-13. Big Sky, MT: IEEE.

Felderhof, L., and D. Gillieson. 2011. "Near-infrared Imagery from Unmanned Aerial Systems and Satellites Can Be Used to Specify Fertilizer Application Rates in Tree Crops." *Canadian Journal of Remote Sensing* 37 (4): 376-386. doi: 10.5589/m11-046.

Firpo, G., R. Salvini, M. Francioni, and P. G. Ranjith. 2011. "Use of Digital Terrestrial Photogrammetry in Rocky Slope Stability Analysis by Distinct Elements Numerical Methods." *International Journal of Rock Mechanics and Mining Sciences* 48 (7): 1045-1054. doi: 10.1016/j.ijrmms.2011.07.007.

Fletcher, A. T., and P. D. Erskine. 2012. "Mapping of a Rare Plant Species (Boronia Deanei) Using Hyper-Resolution Remote Sensing and Concurrent Ground Observation." *Ecological Management & Restoration* 13 (2): 195-198. doi: 10.1111/j.1442-8903.2012.00649.x.

Furfaro, R., B. D. Ganapol, L. F. Johnson, and S. Herwitz. 2005. "Model-Based Neural Network Algorithm for Coffee Ripeness Prediction Using Helios UAV Aerial Images." In *SPIE Proceedings of Remote Sensing for Agriculture, Ecosystems and Hydrology*, 59760X-1-59760X-11. Bruges: SPIE.

Garcia-Ruiz, F., S. Sankaran, J. M. Maja, W. S. Lee, J. Rasmussen, and R. Ehsani. 2013. "Comparison of Two Aerial Imaging Platforms for Identification of Huanglongbing-Infected Citrus Trees." *Computers and Electronics in Agriculture* 91 (Feb.): 106-115. doi: 10.1016/j.compag.2012.12.002.

Gay, A., T. Stewart, R. Angel, M. Easey, A. Eves, N. Thomas, and A. Kemp. 2009. "Developing Unmanned Aerial Vehicles for Local and Flexible Environmental and Agricultural Monitoring." In *Proceedings of the Remote Sensing and Photogrammetry Society Conference*, 471-476. Leicester: ISPRS.

Gertler, J. 2012. *US Unmanned Aerial Systems*, *Report R42136*. Washington: Library of Congress Washington DC Congressional Research Service.

Gillespie, T. W., J. Chu, E. Frankenberg, and D. Thomas. 2007. "Assessment and Prediction of Natural Hazards from Satellite Imagery." *Progress in Physical Geography* 31 (5): 459-470. doi: 10.1177/0309133307083296.

Gimenez, R., I. Marzolff, M. A. Campo, M. Seeger, J. B. Ries, J. Casali, and J. Alvarez-Mozos. 2009. "Accuracy of High-Resolution Photogrammetric Measurements of Gullies with Contrasting Morphology." *Earth Surface Processes and Landforms* 34 (14): 1915-1926. doi: 10.1002/esp.1868.

Gini, R., D. Passoni, L. Pinto, G. Sona. 2012. "Aerial Images from an UAV System: 3D Modeling and Tree Species Classification in a Park Area." In *Proceedings of XXII ISPRS Congress*, 361-366. Melbourne: ISPRS.

Goktogan, A. H., S. Sukkarieh, M. Bryson, J. Randle, T. Lupton, and C. Hung. 2010. "A Rotary-Wing Unmanned Air Vehicle for Aquatic Weed Surveillance and Management." *Journal of Intelligent and Robotic Systems* 57: 467-484. doi: 10.1007/s10846-009-9371-5.

Gong, J., Y. Yue, J. Zhu, Y. Wen, Y. Li, J. Zhou, D. Wang, C. Yu. 2012. "Impacts of the Wenchuan Earthquake on the Chaping River Upstream Channel Change." *International Journal Remote Sensing* 33 (12): 3907-3929. doi: 10.1080/01431161.2011.636767.

Grenzdorffer, G. J., and F. Niemeyer. 2011. "UAV based BRDF-measurements of Agricultural Surfaces with PFIFFIKUS." In *Proceedings of ISPRS Conference on Unmanned Aerial Vehicle in Geomatics*, 1-6. Zurich: ISPRS.

Gruen, A. 2000. "Potential and Limitations of High-resolution Satellite Imagery." In *Proceedings of 21st Asian Conference on Remote Sensing*. Taipei.

Guillen-Climent, M. L., P. J. Zarco-Tejada, J. A. J. Berni, P. R. J. North, F. J. Villalobos. 2012. "Mapping Radiation Interception in Row-Structured Orchards Using 3D Simulation and High-Resolution Airborne Imagery Acquired from a UAV." *Precision Agriculture* 13 (4): 473-500. doi: 10.1007/s11119-012-9263-8.

Habib, A., A. Pullivelli, E. Mitishita, M. Ghanma, and E. M. Kim. 2006. "Stability Analysis of Low-Cost Digital Cameras for Aerial Mapping Using Different Georeferencing Techniques." *The Photogrammetric Record* 21 (113): 29-43. doi: 10.1111/j.1477-9730.2006.00352.x.

Hakala, T., J. Suomalainen, and J. I. Peltoniemi. 2010. "Acquisition of Bidirectional Reflectance Factor Dataset Using a Micro Unmanned Aerial Vehicle and a Consumer Camera." *Remote Sensing* 2 (3): 819-832. doi: 10.3390/rs2030819.

Han, Y. 2009. "An Autonomous Unmanned Aerial Vehicle-Based Imagery System Development and Remote Sensing Images Classification for Agricultural Applications." Master of Science Thesis, Utah State University.

Han-Ya, I., K. Ishii, and N. Noguchi. 2010. "Satellite and Aerial Remote Sensing for Production Estimates and Crop Assessment." *Environmental Control in Biology* 48 (2): 51-58. doi: 10.2525/ecb.48.51.

Hardin, P. J., and T. J. Hardin. 2010. "Small-Scale Remotely Piloted Vehicles in Environmental Research." *Geography Compass* 4 (9): 1297-1311. doi: 10.1111/j.1749-8198.2010.00381.x.

Hardin, P. J., and R. R. Jensen. 2011. "Small-Scale Unmanned Aerial Vehicles in Environmental Remote Sensing: Challenges and Opportunities." *GIScience & Remote Sensing* 48 (1): 99-111. doi: 10.2747/1548-1603.48.1.99.

Hermosilla, T., E. Bermejo, A. Balaguer, and L. A. Ruiz. 2008. "Non-Linear Fourth-Order Image Interpolation for Subpixel Edge Detection and Localization." *Image and vision computing* 26 (9): 1240-1248. doi: 10.1016/j.imavis. 2008.02.012.

Hernández López, D., B. F. García, J. G. Piqueras, and G. V. Alcázar. 2011. "An Approach to the Radiometric Aerotriangulation of Photogrammetric Images." *ISPRS Journal of Photogrammetry and Remote Sensing*, 66 (6): 883-893. doi: 10.1016/j.isprsjprs.2011.09.011.

Herwitz, S. R., L. F. Johnson, J. Arvesen, R. Higgins, J. Leung, and S. Dunagan. 2002. "Precision Agriculture as a Commercial Application for Solar-Powered Unmanned Aerial Vehicles." In

Proceedings of the American Institute of Aeronautics and Astronautics UAV Conference, 1-7. Portsmouth, VA: AIAA.

Hodgson, A. J., M. Noad, H. Marsh, J. Lanyon, and E. Kniest. 2010. *Using Unmanned Aerial Vehicles for Surveys of Marine Mammals in Australia: Test of Concept; Final Report to the Australian Marine Mammal Centre*. Kingston: Australian Government.

Honkavaara, E., T. Hakala, L. Markelin, T. Rosnell, H. Saari, and J. Makynen. 2012. "A Process for Radiometric Correction of UAV Image Blocks." *Photogrammetrie-Fernerkundung-Geoinformation*, 2012 (2): 115-127. doi: 10.1127/1432-8364/2012/0106.

Honkavaara, E., J. Kaivosoja, J. Mäkynen, I. Pellikka, L. Pesonen, H. Saari, H. Salo, T. Hakala, L. Marklelin, and T. Rosnell. 2012. "Hyperspectral Reflectance Signatures and Point Clouds for Precision Agriculture by Light Weight UAV Imaging System." In *Proceedings of XXII ISPRS Congress*, 353-358. Melbourne: ISPRS.

Honkavaara, E., H. Saari, J. Kaivosoja, I. Pölönen, T. Hakala, P. Litkey, J. Mäkynen L. Pesonen. 2013. "Processing and Assessment of Spectrometric, Stereoscopic Imagery Collected Using a Lightweight UAV Spectral Camera for Precision Agriculture." *Remote Sensing* 5(10): 5006-5039. doi: 10.3390/rs5105006.

Hruska, R., J. Mitchell, M. Anderson, and N. F. Glenn. 2012. "Radiometric and Geometric Analysis of Hyperspectral Imagery Acquired from an Unmanned Aerial Vehicle." *Remote Sensing* 4(9): 2736-2752. doi: 10.3390/rs4092736.

Hunt, J. E. R., M. Cavigelli, C. S. T. Daughtry, J. E. Mcmurtrey, and C. L. Walthall. 2005. "Evaluation of Digital Photography from Model Aircraft for Remote Sensing of Crop Biomass and Nitrogen Status." *Precision Agriculture* 6 (4): 359-378. doi: 10.1007/s11119-005-2324-5.

Hunt, J. E. R., W. D. Hively, S. J. Fujikawa, D. S. Linden, C. S. T. Daughtry, and G. W. McCarty. 2010. "Acquisition of NIR-Green-Blue Digital Photographs from Unmanned Aircraft for Crop Monitoring." *Remote Sensing* 2 (1): 290-305. doi: 10.3390/rs2010290.

Inoue, J., J. A. Curry, and J. A. Maslanik. 2008. "Application of Aerosondes to Melt-Pond Observations Over Arctic Sea Ice." Journal of Atmospheric and Oceanic Technology 25 (2): 327-334. doi: http://dx.doi.org/10.1175/2007JTECHA955.1.

Ishihama, F., Y. Watabe, and H. Oguma. 2012. "Validation of a High-Resolution, Remotely Operated Aerial Remote-Sensing System for the Identification of Herbaceous Plant Species." *Applied Vegetation Science* 15 (3): 383-389. doi: 10.1111/j.1654-109X.2012.01184.x.

Israel, M. 2011. "A UAV-Based Roe Deer Fawn Detection System." In *Proceedings of ISPRS Conference on Unmanned Aerial Vehicle in Geomatics*. Zurich, Switzerland.

Jensen, A. M., B. T. Neilson, M. McKee, and Y. Chen. 2012. "Thermal Remote Sensing with an Autonomous Unmanned Aerial Remote Sensing Platform for Surface Stream Temperatures." In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 5049-5052. Munich, IEEE.

Jensen, T., A. Apan, F. Young, and L. Zeller. 2007. "Detecting the Attributes of a Wheat Crop Using Digital Imagery Acquired from a Low-Altitude Platform." *Computers and Electronics in Agriculture* 59 (1-2): 66-77. doi: 10.1016/j.compag.2007.05.004.

Johnson, L. F., S. Herwitz, S. Dunagan, B. Lobitz, D. Sullivan, R. Slye. 2003. "Collection of Ultra High Spatial and Spectral Resolution Image Data over California Vineyards with a Small UAV." In *Proceedings of the 30th International Symposium on Remote Sensing of Environment*, 845-849. Honolulu, HI.

Jones, G. P., IV. 2003. "The Feasibility of Using Small Unmanned Aerial Vehicles for Wildlife Research." Master of Science Thesis, University of Florida.

Jones, G. P., IV., L.G. Pearlstine, H. F. Percival. 2006. "An Assessment of Small Unmanned Aerial Vehicles for Wildlife Research." *Wildlife Society Bulletin* 34 (3): 750-758. doi: 10.2193/0091-7648(2006)34[750:AAOSUA]2.0.CO;2.

Joyce, K. E., S. E. Belliss, S. V. Samsonov, S. J. McNeill, and P. J. Glassey. 2009. "A Review of the Status of Satellite Remote Sensing and Image Processing Techniques for Mapping Natural Hazards and Disasters." *Progress in Physical Geography* 33 (2): 183-207. doi: 10.1177/0309133309339563.

Jezek, K. C., and R. G. Onstott. 1999. "The Role of Remote Sensing in the Environmental Monitoring of Antarctica." *Polar Geography* 23 (1): 55-70. doi: 10.1080/10889379909377664.

Khorram, S., F. H. Koch, C. F. van der Wiele, and S. A. C. Nelson. 2012. *Remote Sensing*. New York: Springer Briefs in Space Development.

Kim, S., S. Gooa, Y. Park, J. Choi, and M. Choa. 2012. "Strategic Multiple Sensor Data Fusion for Time-Critical Natural Disaster Response." In *Proceedings of XXII ISPRS Congress*, 327-330. Melbourne: ISPRS.

Knoth, C., B. Klein, T. Prinz, and T. Kleinebecker. 2013. "Unmanned Aerial Vehicles as Innovative Remote Sensing Platforms for High-Resolution Infrared Imagery to Support Restoration Monitoring in Cut-Over Bogs." *Applied Vegetation Science* 16 (3): 509–517. doi: 10.1111/avsc.12024.

Koh, L. P., and S. A. Wich. 2012. "Dawn of Drone Ecology: Low-Cost Autonomous Aerial Vehicles for Conservation." *Tropical Conservation Science* 5 (2): 121-132.

Koo, V. C., Y. K. Chan, G. Vetharatnam, M. Y. Chua, C. H. Lim, C. S. Lim, C. C. Thum, et al. 2012. "A New Unmanned Aerial Vehicle Synthetic Aperture Radar for Environmental Monitoring." *Progress in Electromagnetics Research* 122: 245-268. doi: 10.2528/PIER11092604.

Koski, W. R., T. Allen, D. Irelan, G. Buck, P. R. Smith, M. Macrander, M. A. Halick, C. Rushing, D. J. Sliwa, and T. L. McDonald. 2009. "Evaluation of an Unmanned Airborne System for Monitoring Marine Mammals." *Aquatic Mammals* 35 (3): 347-357. doi: 10.1578/AM.35.3.2009.347.

Laliberte, A. S., M. A. Goforth, C. M. Steele, A. Rango. 2011. "Multispectral Remote Sensing from Unmanned Aircraft: Image Processing Workflows and Applications for Rangeland Environments." *Remote Sensing* 3 (11): 2529-2551. doi: 10.3390/rs3112529.

Laliberte, A. S., A. Rango. 2009. "Texture and Scale in Object-based Analysis of Subdecimeter Resolution Unmanned Aerial Vehicle (UAV) Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 47 (3): 761-770. doi: 10.1109/TGRS.2008.2009355.

Laliberte, A. S., A. Rango. 2011. "Image Processing and Classification Procedures for Analysis of Sub-Decimeter Imagery Acquired with an Unmanned Aircraft over Arid Rangelands." *GIScience & Remote Sensing* 48 (1): 4-23. doi: 10.2747/1548-1603.48.1.4.

Lechner, A. M., A. Fletchera, K. Johansen, and P. Erskinea. 2012. “Characterising Upland Swamps Using Object-Based Classification Methods and Hyper-Spatial Resolution Imagery Derived from an Unmanned Aerial Vehicle.” In *Proceedings of XXII ISPRS Congress*, 101-106. Melbourne: ISPRS.

Legleiter, C. J., D. A. Roberts, W. A. Marcus, and M. A. Fonstad. 2004. “Passive Optical Remote Sensing of River Channel Morphology and In-stream Habitat: Physical Basis and Feasibility.” *Remote Sensing of Environment* 93 (4): 493-510. doi: 10.1016/j.rse.2004.07.019.

Lejot, J.; C. Delacourt, H. Piegay, T. Fournier, M. Tremelo, and P. Allemand. 2007. “Very High Spatial Resolution Imagery for Channel Bathymetry and Topography from an Unmanned Mapping Controlled Platform.” Earth Surface Processes and Landforms 32 (11): 1705-1725. doi: 10.1002/esp.1595.

Lelong, C. C. D., P. Burger, G. Jubelin, B. Roux, S. Labbé, and F. Baret. 2008. “Assessment of Unmanned Aerial Vehicles Imagery for Quantitative Monitoring of Wheat Crop in Small Plots.” *Sensors* 8 (5): 3557-3585. doi: 10.3390/s8053557.

Lettang, F. J., R. I. Crocker, W. J. Emery, and J. A. Maslanik. 2013. “Estimating the Extent of Drained Supraglacial Lakes on the Greenland Ice Sheet.” International Journal of Remote Sensing 34 (13): 4754-4768. doi: 10.1080/ 01431161.2013.782118.

Li, C., G. Zhang, T. Lei, and A. Gong. 2011. “Quick Image-Processing Method of UAV without Control Points Data in Earthquake Disaster Area.” *Transactions of Nonferrous Metals Society of China* 21: s523-s528. doi: 10.1016/S1003-6326(12)61635-5.

Li, N., D. Zhou, F. Duan, S. Wang, and Y. Cui. 2010. “Application of Unmanned Airship Image System and Processing Techniques for Identifying of Fresh Water Wetlands at a Community Scale.” In *Proceedings of IEEE 18th Geoinformatics International Conference*. Beijing: IEEE.

Lin, J. 2008. “Ice Surface Topography Digital Elevation Model by Interferometric SAR Method.” *GIScience & Remote Sensing* 45 (3): 306-329. doi: 10.2747/1548-1603.45.3.306.

Lin, J., H. Tao, Y. Wang, and Z. Huang. 2010. “Practical Application of Unmanned Aerial Vehicles for Mountain Hazards Survey.” In *Proceedings of IEEE 18th Geoinformatics International Conference*. Beijing: IEEE.

Lin, Z., J. Jiang, and Z. Wang. 2011. “Edge Detection in the Feature Space.” *Image and Vision Computing* 29 (2): 142-154. doi: 10.1016/j.imavis.2010.08.008.

Lubin, D., and R. Massom. 2007. “Remote Sensing of Earth's Polar Regions: Opportunities for Computational Science.” *Computing in Science & Engineering* 9 (1): 58-71. doi: 10.1109/MCSE.2007.16.

Lucieer, A., S. A. Robinson, and D. Turner. 2010. “Using an Unmanned Aerial Vehicle (UAV) for Ultra-High Resolution Mapping of Antarctic Moss Beds.” In *Proceedings of 15th Australasian Remote Sensing & Photogrammetry Conference*. Alice Springs, Australia.

Mackenzie, D. I. 2005. “What Are the Issues with Presence-absence Data for Wildlife Managers?” *Journal of Wildlife Management* 69 (3): 849-860. doi: 10.2193/0022-541X(2005)069[0849:WATIWP]2.0.CO;2.

Malos, J., B. Beamish, L. Munday, P. Reid, and C. James. 2013. “Remote Monitoring of Subsurface Heatings in Opencut Coal Mines.” In *Proceedings of 13th Coal Operators' Conference*, 227-231. Wollongong, Australia.

Martinez, D. J. R., L. Merino, F. Caballero, and A. Ollero. 2011. “Automatic Forest-Fire Measuring Using Ground Stations and Unmanned Aerial Systems.” *Sensors* 11 (6): 6328-6353. doi: 10.3390/s110606328.

McGwire, K. C., M. A. Weltz, J. A. Finzel, C. E. Morris, L. F. Fenstermaker, and D. S. McGraw. 2013. “Multiscale Assessment of Green Leaf Cover in a Semi-Arid Rangeland with a Small Unmanned Aerial Vehicle.” *International Journal of Remote Sensing* 34 (5): 1615-1632. doi: 10.1080/01431161.2012.723836.

Merino, L., F. Caballero, J. R. Martinez-de Dios, J. Ferruz, and A. Ollero. 2006. “A Cooperative Perception System for Multiple UAVs: Application to Automatic Detection of Forest Fires.” *Journal of Field Robotics* 23 (3-4): 165-184. doi: 10.1002/rob.20108.

Mostafa, M. M., and K. P. Schwarz. 2001. “Digital Image Georeferencing from a Multiple Camera System by GPS/INS.” *ISPRS Journal of Photogrammetry and Remote Sensing* 56 (1): 1-12. doi: 10.1016/S0924-2716(01)00030-2.

Murakami, T., M. Yui, and K. Amaha. 2012. “Canopy Height Measurement by Photogrammetric Analysis of Aerial Images: Application to Buckwheat (Fagopyrum Esculentum Moench) Lodging Evaluation.” *Computers and Electronics in Agriculture* 89 (Nov.): 70-75. doi: 10.1016/j.compag. 2012.08.003.

Niethammer, U., M. R. James, S. Rothmund, J. Travelletti, and M. Joswig. 2012. “UAV-based Remote Sensing of the Super-Sauze Landslide: Evaluation and Results.” *Engineering Geology* 128 (March): 2-11. doi: 10.1016/j. enggeo.2011.03.012.

d'Oleire-Oltmanns, S., I. Marzolff, K. D. Peter, and J. B. Ries. 2012. “Unmanned Aerial Vehicle (UAV) for Monitoring Soil Erosion in Morocco.” *Remote Sensing* 4 (11): 3390-3416. doi: 10.3390/rs4113390.

Ollero, A., J. R. Martínez-de-Dios, and L. Merino. 2006. “Unmanned Aerial Vehicles as Tools for Forest-Fire Fighting.” In *Proceedings of V International Conference on Forest Fire Research*, edited by Domingos Xavier Viegas. Coimbra, Portugal.

Pastor, E., C. Barrado, P. Royo, E. Santamaria, J. Lopez, and E. Salami. 2011. “Architecture for a Helicopter-Based Unmanned Aerial Systems Wildfire Surveillance System.” *Geocarto International* 26 (2): 113-131. doi: 10.1080/10106049.2010.531769.

Primicerio, J., S. F. Di Gennaro, E. Fiorillo, L. Genesio, E. Lugato, A. Matese, and F. P. Vaccari. 2010. “A Flexible Unmanned Aerial Vehicle for Precision Agriculture.” *Precision Agriculture* 13 (4): 517-523. doi: 10.1007/s11119-012-9257-6.

Qian, Y., C. Shengbo, L. Peng, C. Tengfei, M. Ming, L. Yanli, Z. Chao, and Z. Liang. 2012. “Application of Low-Altitude Remote Sensing Image by Unmanned Airship in Geological Hazards Investigation.” In *Proceedings of Image and Signal Processing 5th International Congress*, 1015-1018. Chongqing: IEEE.

Rango, A., and A. S. Laliberte. 2012. “Impact of Flight Regulations on Effective Use of Unmanned Aircraft Systems for Natural Resources Applications.” *Journal of Applied Remote Sensing* 4 (1): 1-12. doi:10.1117/1.3474649.

Rathinam, S., P. Almeida, Z. Kim, S. Jackson, A. Tinka, W. Grossman, and R. Sengupta. 2007. “Autonomous Searching and Tracking of a River Using an UAV.” In *Proceedings of American Control Conference*, 359-364, New York: IEEE.

Rau, Y. J., J. P. Jhan, F. Lob, Y. S. Linb. 2011. “Landslide Mapping Using Imagery Acquired by a Fixed-wing UAV.” In *Proceedings of ISPRS Conference on Unmanned Aerial Vehicle in Geomatics*. Zurich: ISPRS.

Reid, A., F. Ramos, and S. Sukkarieh. 2011. “Multi-class Classification of Vegetation in Natural Environments Using an Unmanned Aerial System.” In *Proceedings of IEEE International Conference on Robotics and Automation*, 2953-2959. Shanghai: IEEE.

Samseemoung, G., P. Soni, H. P. Jayasuriya, and V. M. Salokhe. 2012. “Application of Low Altitude Remote Sensing (LARS) Platform for Monitoring Crop Growth and Weed Infestation in a Soybean Plantation.” *Precision Agriculture* 13 (6): 611-627. doi: 10.1007/s11119-012-9271-8.

Sanyal, J., and X. X. Lu. 2004. “Application of Remote Sensing in Flood Management with Special Reference to Monsoon Asia: A Review.” *Natural Hazards* 33 (2): 283-301. doi: 10.1023/B:NHAZ.0000037035.65105.95.

Sarda-palomera, F., G. Bota, C. Viñolo, O. Pallarés, V. Sazatornil, L. Brotons, S. Gomáriz, and F. Sardà. 2012. “Fine-Scale Bird Monitoring from Light Unmanned Aircraft Systems.” *Ibis* 154 (1): 177-183. doi: 10.1111/j.1474-919X.2011.01177.x.

Schoonmaker, J., T. Wells, G. Gilbert, Y. Podobna, I. Petrosyuk, and J. Dirbas. 2008. “Spectral Detection and Monitoring of Marine Mammals.” In *Proceedings of SPIE Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications*, 694606-1- 694606-1-9. Orlando, FL: SPIE.

Schueller, J. K., W. S. Lee, C. D. Crane, and L. R. Parson. 2006. “Remotely-Piloted Helicopter Citrus Yield Map Estimation.” In *Proceedings of ASABE Annual International Meeting*, 1-11. Portland, OR: ASABE.

Smith, J. G., J. Dehn, R. P. Hoblitt, R. G. LaHusen, J. B. Lowenstern, S. C. Moran, L. McClelland, et al. 2009. “Volcano Monitoring.” In *Geological Monitoring*, edited by Young, R., and Norby, L., 273–305. Boulder, Co: Geological Society of America.

Stagakis, S., V. Gonzalez-Dugo, P. Cid, M. L. Guillen-Climent, and P. J. Zarco-Tejada. 2012. “Monitoring Water Stress and Fruit Quality in an Orange Orchard under Regulated Deficit Irrigation Using Narrow-Band Structural and Physiological Remote Sensing Indices.” *ISPRS Journal of Photogrammetry and Remote Sensing* 71: 47-61. doi: 10.1016/j.isprsjprs.2012.05.003.

Stine, R. S., D. Chaudhuri, P. Ray, P. Pathak, and M. Hall-Brown. 2010. “Comparison of Digital Image Processing Techniques for Classifying Arctic Tundra.” *GIScience & Remote Sensing* 47 (1): 78-98. doi: 10.2747/1548-1603.47.1.78.

Strecha, C., A. Fletcher, A. Lechner, P. Erskine, and P. Fua. 2012. “Developing Species Specific Vegetation Maps Using Multi-Spectral Hyperspatial Imagery from Unmanned Aerial Vehicles.” In *Proceedings of XXII ISPRS* Congress: 311-316. Melbourne: ISPRS.

Stumpf, A., T. A. Lampert, J. Malet, and N. Kerle. 2012. “Multi-Scale Line Detection for Landslide Fissure Mapping.” In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 5450-5453. Munich: IEEE.

Sugiura, R., N. Noguchi, and K. Ishii. 2005. “Remote-Sensing Technology for Vegetation Monitoring Using an Unmanned Helicopter.” *Biosystems Engineering* 90 (4): 369-379. doi: 10.1016/j.biosystemseng.2004.12.011.

Suzuki, T., Y. Amano, T. Hashizume, S. Suzuki, and A. Yamaba. 2010. “Generation of Large Mosaic Images for Vegetation Monitoring Using a Small Unmanned Aerial Vehicle.” *Journal of Robotics and Mechatronics* 22 (2): 212-213.

Swain, K. C., S. J. Thomson, and H. P. W. Jayasuriya. 2010. “Adoption of an Unmanned Helicopter for Low-Altitude Remote Sensing to Estimate Yield and Total Biomass of a Rice Crop.” *Transactions of the ASABE* 53 (1): 21-27.

Syvitski, J. P. M., I. Overeem, G. R. Brakenridge, and M. Hannon. 2012. “Floods, Floodplains, Delta Plains - A Satellite Imaging Approach.” *Sedimentary Geology* 267-268: 1–14. doi: 10.1016/j.sedgeo.2012.05.014.

Tahar, K. N., A. Ahmad, and W. Akib. 2011. “Unmanned Aerial Vehicle Technology for Low Cost Landslide Mapping.” In *Proceedings of 11th South East Asian Survey Congress and 13th International Surveyors’ Congress*, 22-31. Kuala Lumpur: PWTC.

Tahar, K. N., A. Ahmad, W. Akib, and W. Mohd. 2012. “A New Approach on Production of Slope Map Using Autonomous Unmanned Aerial Vehicle.” *International Journal of Physical Sciences* 7 (42): 5678-5686.

Tralli, D. M., R. G. Blom, V. Zlotnicki, A. Donnellan, and D. L. Evans. 2005. “Satellite Remote Sensing of Earthquake, Volcano, Flood, Landslide and Coastal Inundation Hazards.” *ISPRS Journal of Photogrammetry and Remote Sensing* 59 (4): 185-198. doi: 10.1016/j.isprsjprs.2005.02.002.

Turner, D., A. Lucieer, and C. Watson. 2011. “Development of an Unmanned Aerial Vehicle (UAV) for Hyper Resolution Vineyard Mapping Based on Visible, Multispectral, and Thermal Imagery.” In *Proceedings of 34th International Symposium on Remote Sensing of Environment*, 4-7.Sydney.

Turner, D., A. Lucieer, and C. Watson. 2012. “An Automated Technique For Generating Georectified Mosaics From Ultra-High Resolution Unmanned Aerial Vehicle (UAV) Imagery, Based On Structure From Motion (SFM) Point Clouds.” *Remote Sensing* 4 (5): 1392-1410. doi: 10.3390/rs4051392.

Uto, K., H. Seki, G. Saito, and Y. Kosugi. 2013. “Characterization of Rice Paddies by a UAV-Mounted Miniature Hyperspectral Sensor System.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (2): 851- 860. doi: 10.1109/JSTARS.2013.2250921.

Valavanis, K. P. 2007. *Advances in Unmanned Aerial Vehicles: State of The Art and the Road to Autonomy*. Dordrecht: Springer.

Van Persie, M., A. Oostdijk, J. Fix, M. C. Van Sijl, and L. Edgardh. 2011. “Real-Time UAV based Geospatial Video Integrated into the Fire Brigades Crisis Management GIS System.” In *Proceedings of ISPRS Conference on Unmanned Aerial Vehicle in Geomatics*, 173-175. Zurich: ISPRS.

Vermeulen, C., P. Lejeune, J. Lisein, P. Sawadogo, and P. Bouche. 2013. “Unmanned Aerial Survey of Elephants.” *PLOS One* 8 (2): e54700-1- e54700-7. doi: 10.1371/journal.pone.0054700.

Vogler, A., H. Eisenbeiss, I. Aulinger-Leipner, and P. Stamp. 2009. “Impact of Topography on Cross-Pollination in Maize (Zea Mays L.).” *European Journal of Agronomy* 31 (2): 99-102. doi: 10.1016/j.eja.2009.04.003.

Watts, A. C., V. G. Ambrosia, and E. A. Hinkley. 2012. “Unmanned Aircraft Systems in Remote Sensing and Scientific Research: Classification and Considerations of Use.” *Remote Sensing* 4 (6): 1671-1692. doi:10.3390/rs4061671.

Wawrzyniak, V., H. Piegay, P. Allemand, L. Vaudor, and P. Grandjean. 2013. “Prediction of Water Temperature Heterogeneity of Braided Rivers Using Very High Resolution Thermal Infrared (TIR) Images.” *International Journal of Remote Sensing* 34 (13): 4812-4831. doi: 10.1080/01431161.2013.782113.

Wu, J., and G. Zhou. 2006. “Real-Time UAV Video Processing for Quick-Response to Natural Disaster.” In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 976-979. Denver, CO: USA.

Wundram, D., and J. Loffler. 2008. “High-Resolution Spatial Analysis of Mountain Landscapes Using a Low-Altitude Remote Sensing Approach.” *International Journal of Remote Sensing* 29 (4): 961-974. doi: 10.1080/01431160701352113.

Xiang, H. 2008. “Autonomous Agricultural Remote Sensing Systems with High Spatial and Temporal Resolutions.” PhD diss., University of Illinois at Urbana-Champaign.

Xiang, H., and L. Tian. 2011a. “Development of a Low-Cost Agricultural Remote Sensing System based on an Autonomous Unmanned Aerial Vehicle (UAV).” *Biosystems Engineering* 2011a, 108 (2): 174-190. doi: 10.1016/j. biosystemseng. 2010.11.010.

Xiang, H., and L. Tian. 2011b. “Method for Automatic Georeferencing Aerial Remote Sensing (RS) Images from an Unmanned Aerial Vehicle (UAV) Platform.” *Biosystems Engineering* 108 (2): 104-113. doi: 10.1016/j.biosystemseng. 2010.11.003.

Yilmaz, K. K., R. F. Adlerab, Y. Tianbc, Y. Hongd, and H. F. Piercebe. 2010. “Evaluation of a Satellite-Based Global Flood Monitoring System.” *International Journal of Remote Sensing* 31 (14): 3763-3782. doi: 10.1080/01431161.2010.483489.

Yue, J., T. Lei, C. Li, and J. Zhu. 2012. “The Application of Unmanned Aerial Vehicle Remote Sensing in Quickly Monitoring Crop Pests.” *Intelligent Automation & Soft Computing* 18 (8): 1043-1052. doi: 10.1080/10798587.2008.10643309.

Yunxia, H., L. Minzan, Z. Xijie, J. Liangliang, C. Xingping, and Z. Fusuo. 2005. “Precision Management of Winter Wheat Based on Aerial Images and Hyperspectral Data Obtained by Unmanned Aircraft.” In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 3109-3112. Seoul: IEEE.

Zaman, B.; A. M. Jensen, and M. McKee. 2011. “Use of High-Resolution Multispectral Imagery Acquired with an Autonomous Unmanned Aerial Vehicle to Quantify the Spread of an Invasive Wetlands Species.” In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 803-806. Vancouver, BC: IEEE.

Zarco-Tejada, P. J., J. Berni, L. Suarez, G. Sepulcre-Canto, F. Morales, and J. R. Miller. 2009. “Imaging Chlorophyll Fluorescence with an Airborne Narrow-Band Multispectral Camera for Vegetation Stress Detection.” *Remote Sensing of Environment* 113 (6): 1262-1275. doi: 10.1016/j.rse.2009.02.016.

Zarco-Tejada, P. J., V. Gonzalez-Dugo, and J. Berni. 2012. “Fluorescence, Temperature and Narrow-Band Indices Acquired from a UAV Platform for Water Stress Detection Using a Micro-Hyperspectral Imager and a Thermal Camera.” *Remote Sensing of Environment* 117 (Feb.): 322-337. doi: 10.1016/j.rse.2011.10.007.

Zarco-Tejada, P. J., M. L. Guillen-Climent, R. Hernandez-Clemente, A. Catalina, M. R. Gonzalez, and P. Martin. 2013. “Estimating Leaf Carotenoid Content in Vineyards Using High Resolution

Hyperspectral Imagery Acquired from an Unmanned Aerial Vehicle (UAV)." *Agricultural and Forest Meteorology* 171-172: 281-294. doi: 10.1016/j.agrformet.2012.12.013.

Zhang, C., and J. M. Kovacs. 2012. "The Application of Small Unmanned Aerial Systems for Precision Agriculture: A Review." *Precision Agriculture* 13 (6): 693–712. doi: 10.1007/s11119-012-9274-5.

Zhou, G. 2009. "Near real-time Orthorectification and Mosaic of Small UAV Video Flow for Time-critical Event Response." *IEEE Transactions on Geoscience and Remote Sensing* 47 (3): 739-747. doi: 10.1109/TGRS.2008.2006505.

Zhu, J., K. Wang, J. Deng, and T. Harmon. 2009. "Quantifying Nitrogen Status of Rice Using Low Altitude UAV-Mounted System and Object-Oriented Segmentation Methodology." In *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 1-7. San Diego, CA: ASME.

Chapter 3

3. System Development: technical and methodological aspects

3.1 Article Presentation

3.1.1 Background

This article is part of the thesis related to the second specific objective, namely exploring the methodological and experimental aspects of implementing a UAV-PS for ensuring high-quality visual and topographic data. This article was originally published at Sensor Journal, available online: 10.3390/s15112749.

In terms of research problematic, this article considers the requirements of an efficient UAV-PS for precise surveying and volumetric change detection in an open-pit mine. This case study was selected due to its challenging conditions, which can well represent other types of environmental applications as well. These characteristics are discussed in the article. Based on previous studies performed on UAV-photogrammetry systems, this article defines important aspects of system development, data acquisition, and data processing, discusses the issues and challenges of each aspect, proposes solutions to address them, and evaluates the suggested solutions extensively.

Please note that 3D point cloud generation at this article is performed using commercial software. The reason is that this particular part of the thesis does not concern with the challenges of photogrammetric software solutions. Instead, it is concerned with the issues that need to be addressed (at the levels of system development, deployment, acquisition and pre-processing) for producing high-quality data, regardless of the photogrammetric software solution. These issues include the followings.

- Stability of intrinsic camera parameters
  - Accuracy of offline calibration
  - Accuracy of on-the-job self-calibration
- Quality of direct geo-referencing
  - Limits of navigation sensors
  - Quality of platform calibration and sensor integration
- Impacts of imaging-network configuration
  - Scale consistency and amount of overlap between images
  - Importance of planning flight trajectory and camera shots w.r.t. terrain characteristics
- Optimality of indirect geo-referencing
  - Number and distribution of GCPs
  - Positioning accuracy of GCPs
  - Precision of GCP detection on images
- Effects of high image intensity variations such as presence of shadows

3.1.2 Methodology

The flowchart of the research methodology for system development is presented in Figure 3.1.1. The following article discusses these steps in detail.

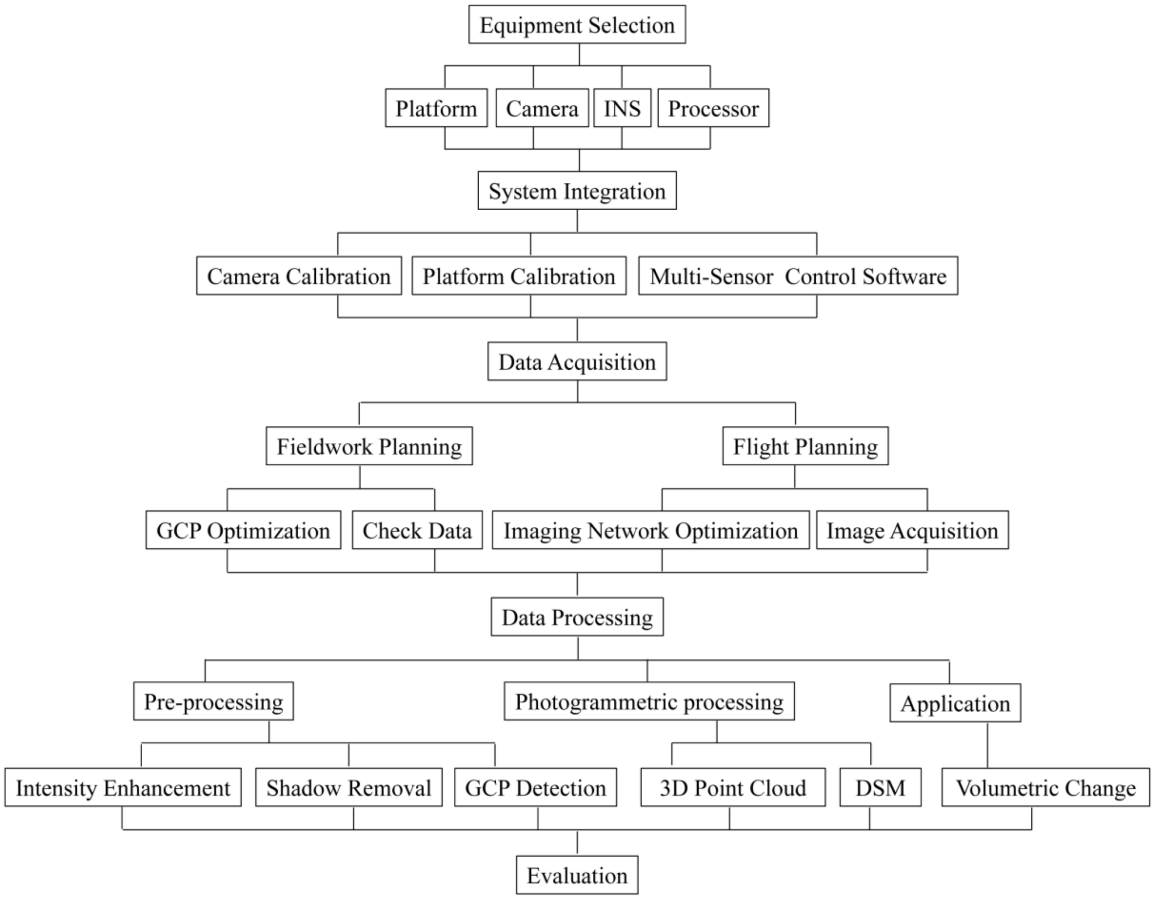


Figure 3.1.1. Methodological flowchart for system development

Development and Evaluation of a UAV-Photogrammetry System for Precise 3D Environmental Modeling

by Mozhdeh Shahbazi, Gunho Sohn, Jérôme Théau and Patrick Ménard

Sensors,

ISSN: 1424-8220

Volume 15, No. 11, October 2015, pp. 27493-27524

Abstract

The specific requirements of UAV-photogrammetry necessitate particular solutions for system development, which have mostly been ignored or not assessed adequately in recent studies. Accordingly, this paper presents the methodological and experimental aspects of correctly implementing a UAV-photogrammetry system. The hardware of the system consists of an electric-powered helicopter, a high-resolution digital camera and an inertial navigation system. The software of the system includes the in-house programs specifically designed for camera calibration, platform calibration, system integration, on-board data acquisition, flight planning and on-the-job self-calibration. The detailed features of the system are discussed, and solutions are proposed in order to enhance the system and its photogrammetric outputs. The developed system is extensively tested for precise modeling of the challenging environment of an open-pit gravel mine. The accuracy of the results is evaluated under various mapping conditions, including direct georeferencing and indirect georeferencing with different numbers, distributions and types of ground control points. Additionally, the effects of imaging configuration and network stability on modeling accuracy are assessed. The experiments demonstrated that 1.55 m horizontal and 3.16 m vertical absolute modeling accuracy could be achieved via direct geo-referencing, which was improved to 0.4 cm and 1.7 cm after indirect geo-referencing.

Resumé

Les besoins spécifiques de la photogrammétrie à l’aide des drones exigent des solutions adaptées pour le développement des systèmes, qui ont souvent été ignorées ou non évaluées de manière adéquate dans les études récentes. Par conséquent, cet article présente les aspects méthodologiques et expérimentaux pour la mise en œuvre d’un système photogrammétrique fonctionnel à l’aide des drones. Le système est constitué d'un hélicoptère à propulsion électrique, d’une caméra numérique à haute résolution et d’une station inertielle. Le logiciel de contrôle comprend des programmes conçus spécifiquement pour la calibration de la caméra, la calibration de la plateforme, l'intégration de système, l'acquisition des données, la planification de vol et l’autocalibration. Les caractéristiques détaillées du système sont discutées et des solutions sont proposées afin d'améliorer les systèmes et les produits photogrammétriques. Le système développé est testé de manière approfondie pour la modélisation précise d’une gravière à ciel ouvert, représentative d’un environnement complexe. La précision des résultats est évaluée dans diverses conditions de cartographie comprenant les géoréférencements direct et indirect incluant différentes quantités, distributions et types de points de contrôle au sol. De plus, les effets de la configuration de l'imagerie et de la stabilité du réseau sur la précision de la modélisation sont évalués. Les expériences ont démontré une précision de la modélisation absolue par l'intermédiaire du géoréférencement direct de 1,55 m horizontal et 3,16 m vertical. Cette précision a été améliorée à 0,4 cm et 1,7 cm après le géoréférencement indirect.

Keywords: UAV; modeling; photogrammetry; calibration; georeferencing; ground control point; mine

3.2 Introduction

3.2.1 Background

Unmanned aerial imagery has recently been applied in various domains such as natural resource management, spatial ecology and civil engineering [1–3]. Most of these unmanned aerial vehicles (UAV) applications require geospatial information of the environment. Consequently, three-dimensional (3D) environmental modeling via UAV-photogrammetry systems (UAV-PS) has become a matter of growing interest among both researchers and industries. However, a surveying-grade UAV-PS has critical differences from traditional photogrammetry systems, which should be considered carefully in its development and application. The following paragraphs discuss the background of UAV-PSs and the efforts made to evaluate their capacities.

Typically, development of a UAV-PS starts with selecting the platform as well as the imaging and navigation sensors compatible with it. Regarding the platform, the payload capacity, endurance, range, degree of autonomy must be considered. In some studies, pre-packaged UAVs are used, e.g., AscTec Falcon8 [4], Aeryon Scout [5], SenseFly eBee [6]. Such systems offer safety and ease of operation. However, they offer less flexibility regarding sensor selection and adjustment.

Navigation sensors play two roles in a UAV-PS: auto-piloting the platform and determining the exterior orientation (EO) parameters of images. High-grade GNSS-based inertial navigation systems (INS) can be used in order to eliminate the requirement for establishing ground control points (GCPs) and to achieve enough spatial accuracy via direct georeferencing (DG) [7]. However, consumer-grade systems are preferred considering the costs and limitations of access to base stations for differential or real-time-kinematic (RTK) global-positioning-system (GPS) [8,9]. In such systems, different strategies might be taken for increasing the positioning accuracy—e.g., replacing poor-quality GPS elevation data with height measurements from a barometric altimeter [10]. Accuracy of DG depends on the performance of INS components and the accuracy of platform calibration. Moreover, the system-integration scheme is important since it controls the synchronization between imaging and navigation sensors. Depending on flight speed and accuracy of INS measurements, the delay between camera exposures and their geo-tags can cause serious positioning drifts [7,11].

When indirect georeferencing is performed, considerable care should be given to several factors such as the accuracy of multi-view image matching, on-the-job self-calibration and GCP positioning. Discussed briefly in few studies [10,12,13], the method used to locate GCPs on the images and the configuration of the GCPs are also important factors in determining the final accuracy of indirect georeferencing. Accordingly, the optimum configuration of GCPs required to achieve a certain level of accuracy is a significant concern in the field of UAV-PS. In most of UAV applications, only a minimum number of GCPs in a special configuration and with a limited positioning accuracy can be established. In order to ensure that the results, based on these conditions, can satisfy the accuracy requirements of the application, it is important to have an *a priori* knowledge of the final accuracy.

In terms of imaging sensors, a high-resolution digital visible camera is the key element for photogrammetric mapping. Despite the benefits of non-metric digital cameras such as low price, light

weight and high resolution, the instability of their lens and sensor mounts is still a concern in unmanned aerial mapping. Therefore, intrinsic camera calibration must be performed to determine the interior orientation (IO) and distortion parameters of the camera. When metric accuracies are required, offline camera calibration is suggested [14]. However, offline calibration parameters change slightly during the flight due to platform vibrations and instability of camera components [15]. A solution to this problem is to calibrate the camera by adding its systematic errors as additional parameters to aerial block bundle adjustment (BBA), which is known as self-calibration. However, inaccuracy of image observations may influence the calibration parameters as they are all adjusted together with completely unknown parameters such as object-space coordinates of tie points and EO parameters [16]. Thus, motion blur and noise, which are inevitably present in unmanned aerial images, affect the accuracy of calibration. Besides, the numerical stability of self-calibration decreases highly depending on the aerial imaging configuration. Therefore, careful solutions are required to address the issues of on-the-job self-calibration for unmanned aerial imagery.

3.2.2 Environmental Application

Regarding the environmental application, the system of this study was applied for gravel-pit surveying and volumetric change measurement. This environment was selected because of two reasons. Firstly, open-pit mines provide a challenging environment for 3D modeling. That is, considerable scale variations are introduced to the images due to the low altitude of platform in comparison with the terrain relief [17]. Secondly, there are several mining and geological applications which require high-resolution accurate 3D information of open-pit mines, e.g., geotechnical risk assessment. Previous studies have shown that the topographic data must provide a ground resolution of 1–3 cm in order to predict hazardous events such as ground subsidence, slope instability and landslides [18]. Furthermore, mining companies have to quantify the amount of extracted mass and stocked material regularly. The map scale required for volumetric measurement in earthworks is usually between 1:4000 and 1:10,000 [19]. Considering the requirements of mining applications, including spatial and temporal resolution, speed of measurement and safety criteria, unmanned aerial systems can be better solutions for mine mapping in comparison with traditional terrestrial surveying techniques. This can be noticed by the significant increase in use of UAV-PSs in mining applications during the last few years [20–22].

1.3. Objectives and Assumptions

This paper presents the details of development and implementation of a UAV-PS. In addition to general aspects of the development, the main focus of this study is to discuss the issues and to perform the experiments that are usually ignored or not thoroughly addressed for UAV-PSs. First, the paper concentrates on the procedures for camera and platform calibration as well as system integration. Instead of discussing the regular aspects of calibration, the main focus is on the design of the test-field and automatic target detection assuming that these elements impact the efficiency of calibration significantly. Regarding the system integration, it is assumed that the developed software solution is able to integrate the navigation and imaging sensors accurately without needing any additional mechanism.

Afterwards, the photogrammetric processing workflow is presented. Some aspects of image pre-processing are discussed and their impacts on the accuracy of modeling are investigated. Then, assuming that the accuracy of on-the-job self-calibration is affected by the imaging network, a BBA strategy is suggested to control this adverse effect. This assumption and efficiency of the BBA strategy are also verified. Furthermore, several experiments are designed to assess the effect of GCPs configuration on modeling accuracy. The main assumption that these experiments verify is that a minimum number of GCPs can provide an accuracy level equivalent to the one achievable with redundant number of GCPs under two conditions. First, they are distributed over the whole zone and their visibility in images is maximized. Second, the imaging configuration is proper. That is the imaging configuration ensures scale consistency of the network.

The rest of the paper is structured as follows: first, the equipment is presented. Then, the procedure of system development, including camera calibration, platform calibration and system integration, are discussed in Sections 3.4. Afterwards, Sections 3.5–3.6 describe the methodology of data acquisition and data processing. The experiments performed to evaluate the system are presented in Section 3.7, and the results are discussed in Section 3.8. At the end, the conclusions and final remarks are presented in Section 3.9.

3.3 Equipment

3.3.1 Platform

The platform used in this project is a Responder helicopter built by ING Robotic Aviation Inc. (Ottawa, ON, Canada) (Figure 3.3.1a). Responder is a vertical take-off & landing UAV which is equipped with a lightweight, carbon-fiber gimbal. This platform has 12 kg payload capacity and cruise operational endurance of 40 min. With our whole independent package of sensors, computer and batteries weighing about 3 kg, the platform could safely fly for 25 min in a day with wind speed of 19 km/h. The platform is equipped with an open-source autopilot—ArduPilot Autopilot Suite. It comes with a portable, compact ground control station to visualize, plan and control autonomous flights (Figure 3.3.1b).

3.3.2 Navigation sensor

The navigation sensor is a GPS-aided INS, MIDGII from Microbotics Inc. (Hampton, VA, USA) (Figure 3.3.1c, stacked on the top of the camera). The unit measures pitch and roll with 0.4° and heading (yaw) with 1–2° of accuracy. Its positioning accuracy is 2–5 m depending on availability of wide area augmentation system (WAAS). The output rate of the unit can be extended up to 50 Hz.

3.3.3 Imaging sensor

The imaging sensor is a GE4900C visible camera (Prosilica, Exton, PA, USA) (Figure 3.3.1c). It has a 36.0528 mm × 24.0352 mm sensor at pixel size of 7.4 μm. It is equipped with a 35 mm F-mount lens and supports minimum exposure time of 625 μs. The fact that the camera has a global shutter and charge-coupled-device (CCD) progressive sensor makes the imaging more robust against motion blur,

interlacing artifact and read-out delay, which are all essential for UAV-PS [23]. Global shutter controls the incoming light all over the image surface simultaneously. Thus, at any time instance, all photo detectors are either equally closed or equally open. This is in contrast with rolling shutters where exposures move row by row from one side to another side of the image. In the CCD architecture, only one-pixel shift happens to move the charge from image to storage area. Therefore, the readout time and energy consumption decrease considerably. Progressive scanning is also strongly preferred for grabbing moving images, since the images are free of interlacing artifacts caused by the time lag of frame fields.

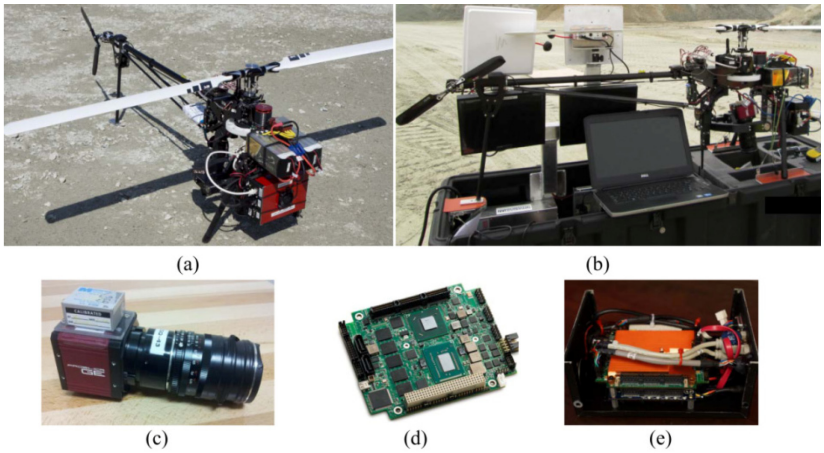


Figure 3.3.1. Equipment (**a**) Aerial platform. (**b**) Ground control station. (**c**) INS attached to camera. (**d**) Computer board. (**e**) Computer board and power supply stacked together.

3.3.4 Onboard computer

The computer applied in this study is an ultra-small, single-board system (CoreModule 920, ADLINK, San Jose, CA, USA), which is based on a 1.7 GHz Intel Core™ i7 processor (Figure 3.3.1d). The board is stacked together with a PC/104 power supply (Figure 3.3.1e). The power supply receives 12.8-volt DC input from a 3200 mAh LiFePO4 battery pack. In return, it provides +5 V regulated DC voltage to the computer, +12 V to the camera and +5 V to a fan for cooling the processing unit. With this configuration, the embedded system is capable of acquiring, logging and storing images with a rate of 3 frames per second and navigation data with a rate of 50 Hz during approximately 70 min.

3.4 System Development

3.4.1 Camera calibration

In this study, offline camera calibration is performed using a test-field (Figure 3.4.1) via a conventional photogrammetric calibration method known as inner-constrained bundle adjustment with additional parameters [24]. In this study, Brown's additional parameters are applied to model the systematic errors of the camera [25]. Therefore, the camera IO parameters, radial and decentering lens distortions as well as in-plane scale and shear distortions of the sensor [26] are modeled via calibration. As digital camera

calibration is a well-studied topic in photogrammetry, detailed theories are avoided here. Instead, other important aspects, including our methodology for test-field design and target detection, are discussed.

Please see Appendix I for more details about the intrinsic camera calibration parameters, the distortion models, the theoretical details of the estimation, and the results of stability analysis.

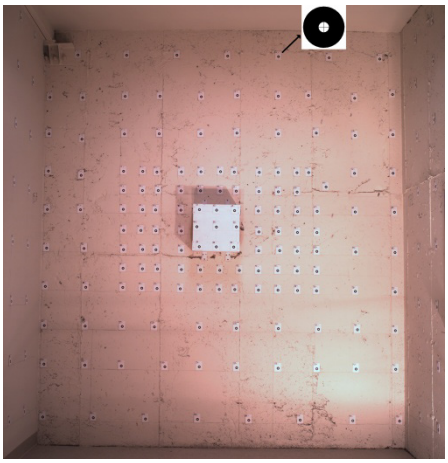


Figure 3.4.1. Camera calibration test-field.

3.4.1.1 Design of the calibration test-field

Two camera parameters determine the size and depth of a test-field: focus distance and field of view. For aerial imaging, camera focus distance is set to infinity so that different altitudes can be covered in the depth of field (DoF) of the camera. However, the focal length extends slightly during focusing (When the focus distance is changed, a small group of elements inside the lens, instead of the whole barrel, is moved to provide the focus. Thus, the focal length changes slightly.). This means that the focus distance should remain fixed all the time. Therefore, it should be ensured that the calibration test-field can provide focused photos at short distances while the focus distance is still set to infinity. Considering Equation (1), the far and near limits of DoF (H_f, H_n) depend on F-number (d), focal length (f), circle of confusion diameter (c) and focus distance (h). That is all the objects located between H_n and H_f from the camera can be imaged sharply:

$$H_f = h \Big/ 1-(h-f)c\frac{d}{f^2}, \qquad H_n = h \Big/ 1+(h-f)c\frac{d}{f^2} \tag{1}$$

If the focus distance is set to infinity $(h \to \infty)$, then the F-number should be increased largely to provide focus at short ranges. By setting the F-number to its maximum value ($d_{\max}$), the minimum focus distance can be determined $(H_{n_{\min}})$. Thus, the distance of the test-field from the camera, namely the test-field depth, should be larger than $H_{n_{\min}}$. Notice that by maximizing the F-number, the aperture opening (A) decreases (Equation (2)), and calibration images become very dark. To compensate this, the exposure time should be increased according to Equation (3). Let t_1 be the minimum exposure time of the camera which is usually selected for aerial imaging to reduce motion blurring artifacts. Accordingly, A_1 is the

aperture opening which provides proper illumination for outdoor acquisition in combination with t_1. If A_2 is the aperture opening when maximizing the F-number, then t_2 is the exposure time that should be set to avoid either underexposure or overexposure:

$$A = \pi f^2 / 4d^2 \tag{2}$$

$$A_1 t_1 = A_2 t_2 \tag{3}$$

Afterwards, the width and height of the test-field should be determined. It is essential to model the systematic errors based on the distortions observed uniformly across the whole image [24]. Therefore, it should be ensured that the test-field is large enough to cover approximately the whole field of view (FoV) of the camera. The horizontal and vertical angles of FoV (α_h, α_v) can be calculated from the sensor size (W, H) and the focal length (f) as in Equation (4). Therefore, the minimum size required for the test-field ($W_t \times H_t$) can be determined via Equation (5)

$$\alpha_h = 2\tan^{-1}(W/2f) \quad , \quad \alpha_v = 2\tan^{-1}(H/2f) \tag{4}$$

$$W_t = 2H_{n_{\min}} \tan(\alpha_h/2) \quad , \quad H_t = 2H_{n_{\min}} \tan(\alpha_v/2) \tag{5}$$

3.4.1.2 Target detection

Figure 3.4.2 demonstrates the approach of this study for detecting the targets. Unless otherwise indicated, the procedures mentioned in the diagram are fully automatic.

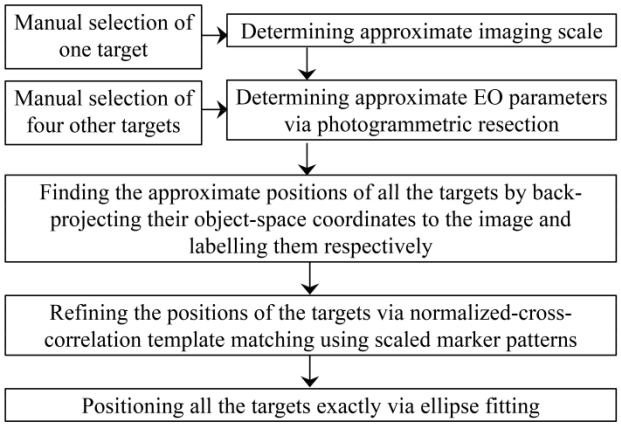


Figure 3.4.2. Diagram of the target detection method.

In this study, the targets are designed as black and white rings with crosshairs passing through the circles' centers. The reason for using circular targets is that once the image of a circle is deformed under any linear transformation, it appears as an ellipse. Then, the techniques of ellipse detection can be applied to position it accurately. The following paragraphs explain the ellipse fitting method developed to determine the accurate positions of the targets.

Assuming that (x_a, y_a) is the approximate position of a target, a rectangular window is centered at (x_a, y_a). The window is transformed to binary format, and its connected black components are detected. Each closed component represents a candidate region. Let B denote the set of all the pixels belonging to the boundary of a candidate region. If the region is actually a target, then its boundary can be modeled as an ellipse with Equation (6):

$$(x-x_o)^2 + K(y-y_o)^2 = R^2 \quad ; \forall (x,y) \in B \tag{6}$$

where (x_o, y_o) is the center, R is the flattened radius and K is the aspect ratio of the ellipse. A small percentage of the points belonging to B are reserved as checkpoints. Other points are served as observations to determine the ellipse parameters with least squares fitting technique. Once the ellipse is defined mathematically, the checkpoints are validated against the ellipse model. If the fitting error is less than a given threshold, then the candidate region is recognized as a target, and the ellipse center denotes the exact position of the target.

3.4.2 Aerial platform calibration

The goal of platform calibration is to make sure that EO parameters of images are represented in an earth-centered, earth-fixed (ECEF) reference system, in which the navigation positioning data are represented as well. To this end, the vector between the perspective center of the camera and the center of the INS body-fixed system—known as lever-arm offset—as well as the rotations of the camera coordinate system with respect to the INS system—known as bore-sight angles—should be determined. In this study, the lever-arm offset is ignored. This is due to the fact that the offset between the INS and the camera never exceeds a few centimeters, which is far below the precision of GPS measurements (a few meters).

Attitude outputs from the INS are presented as Euler angles, also known as Cardan. The Cardan consists of three rotations: roll (ϕ), pitch (θ) and yaw (ψ). The rotation matrix R_b^n —composed of Euler angles— rotates vectors from the INS body-fixed coordinates system (b) to the local geodetic system (n) as in Equation (7). Likewise, the rotation matrix R_n^e —composed of geodetic latitude (φ) and longitude (λ)— rotates vectors from the local geodetic system to the ECEF system (e) as in Equation (8). Therefore, the rotation matrix R_b^e rotates vectors from the INS body-fixed coordinate system to the ECEF system as in Equation (9) [27]:

$$R_b^n = R_z(\psi) R_y(\theta) R_x(\phi) \tag{7}$$

$$R_n^e = R_z(\pi - \lambda) R_y(\pi/2 - \varphi) \tag{8}$$

$$R_b^e = R_n^e R_b^n \tag{9}$$

The required rotation matrix for image georeferencing is R_i^e, which describes the rotations from the camera coordinate system (i) to the ECEF one. The rotation matrix R_i^e can be calculated using the rotation matrix R_b^e and the bore-sight matrix R_i^b :

$$R_i^e = R_b^e R_i^b \tag{10}$$

To determine the bore-sight matrix R_i^b, first, a network of targets is established in the ECEF coordinate system (Figure 3.4.3). Then, the targets are photographed using the camera which is firmly installed on the platform with the INS. Simultaneously, the INS data (R_b^n) is logged. At the post-processing stage, the position ($\vec{X}_i^e$) and orientation (R_i^e) of the camera center are calculated via photogrammetric resection. Using the geodetic coordinates of the camera center $(\varphi_i^e, \lambda_i^e)$—derived from Cartesian coordinates $\vec{X}_i^e$—the rotation matrix R_n^e is calculated. Then, the rotation matrix R_b^e from the INS body-fixed system to the ECEF system is determined via Equation (9). Finally, by substituting R_i^e and R_b^e to Equation (10), the unknown bore-sight matrix R_i^b is determined.

Figure 3.4.3. Test-field for platform calibration.

Notice that, in this study, the camera and the INS were stacked together, in a fixed status as in Figure 3.3.1c. Consequently, the platform could be calibrated before installing the sensors on the UAV.

3.4.3 System integration

A UAV-PS consists of a platform, camera, navigation system and control system. The control system is responsible for various tasks including power control, setting the data-acquisition parameters, data logging, data storage and time synchronization. In this study, the hardware of the control system simply includes the computer and the power supply as described in Section 3.3.4. The software solution developed for this control system contains three main classes: INS, camera, and clock.

The main functionality of the clock class is to get the time up to nanoseconds from the system-wide clock and assign it to any acquisition event in a real-time manner. The INS class is responsible for communication with the INS and recording the navigation messages. Each navigation message contains the information of position, rotation and GPS time. The system time at the moment of receiving each navigation message is also assigned to that message by the clock class. The GPS time of the messages is assigned to a shared variable as well, which has external linkage to the camera class. The camera class is

responsible for communication with the camera and setting its attributes including triggering time interval, exposure and gain value, and frame size. Although several methods of acquisition are available for Prosilica cameras, software triggering mode is used to facilitate the synchronization process. That is the camera is triggered automatically based on defined intervals, e.g., every 500 ms. The end of camera exposure is set as an event, and a callback function is registered to this event. The functionality of this callback is to save the acquired frame and tag the navigation information to it. This information includes the GPS time and the navigation data received from the INS as well as the system time observed at the epoch of the exposure-end event. Finally, the software makes these classes operate together. It starts two threads in the calling process to execute the main functions of the classes simultaneously.

Notice that the GPS time, tagged to each image, is determined by the INS class and the frequency of INS data is 50 Hz. Therefore, the GPS timestamp is theoretically less than 20 ms different from the exact time of the exposure. Assuming a flight speed of 20 km/h, the time-synchronization error of 20 ms can cause 11 cm of shift between the true position and the tagged position of an image. When a navigation-grade GPS is used, such a shift is quite below the precision of GPS measurements and can be ignored [7]. However, if differential GPS measurements are used, then this error must be systematically handled [11]. To do so, in a post-processing step, the difference between the system time and the GPS time that are tagged to each image is used to derive the exact navigation data corresponding to that image via a linear interpolation over the two INS messages, between which the image is acquired.

3.5 Data Acquisition

In this study, the data were acquired from a gravel pit at Sherbrooke, QC, Canada. The extent of the gravel-pit mine is shown in Figure 3.5.1. Two series of data over a period of two months were acquired—August and October 2014. Two main zones were considered for the experiments (Figure 3.5.1). The red zone represents one part of the gravel pit which was covered by stockpiles, and the green zone represents the zone covered by cliffs and rocks.

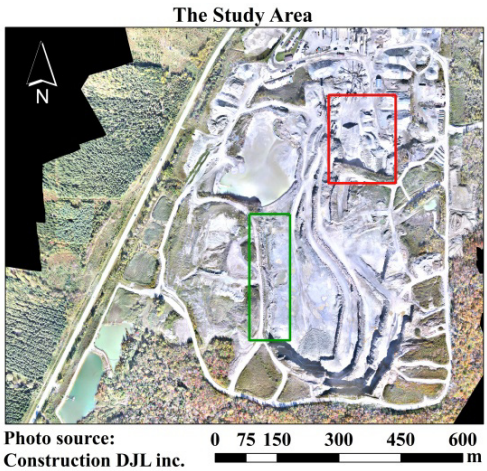


Figure 3.5.1. The study area and mapping zones.

3.5.1 Data-acquisition planning

In order to perform flight planning, there exist several software packages. However, in this study, a simple software solution is developed to satisfy the specific needs of the project for both flight and fieldwork planning. The interface of the software is shown in Figure 3.5.2. The main inputs of the software are the platform and sensor characteristics as well as the desired overlap and ground resolution for imagery. In order to calculate the position of the sun, the flight time is needed too. Knowing the position of the sun helps to minimize shadow effects; the larger the solar elevation angle, the shorter the shadows. The software allows users to either load or graphically choose the predicted positions of GCPs—red triangles in the display panel of Figure 3.5.2. It is, then, possible to determine the flight zone—blue polygon and the flight home—yellow lozenge. The software designs and logs the flight plan afterwards. One of the significant applications of this software is to design the approximate spatial distribution of GCPs considering two conditions. First, GCPs should be installed at stable locations distributed over the whole imaging zone; Second, their visibility in the images should be maximized.

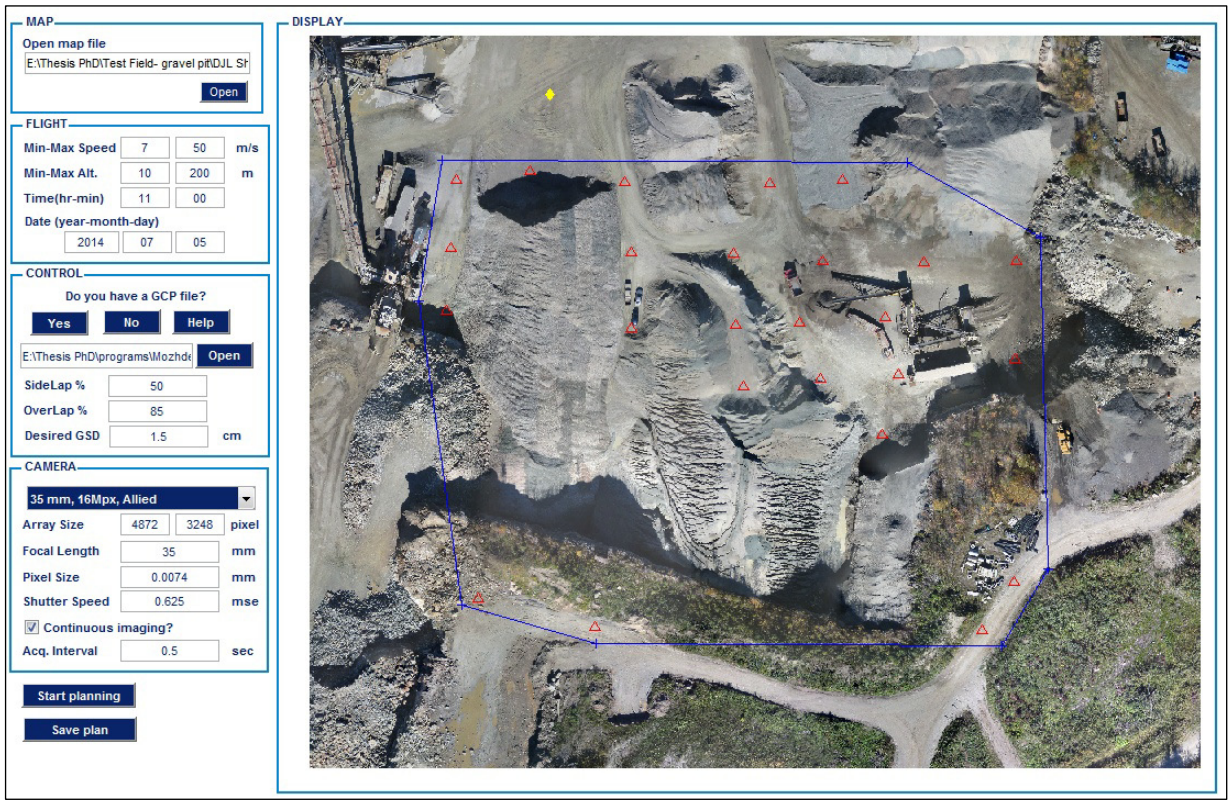


Figure 3.5.2. Interface of the flight-planning software.

3.5.2 Fieldwork

The first task of the fieldwork was to initialize the GPS base receiver for collecting RTK measurements. The absolute coordinates of the base point were determined with 2–5 mm accuracy. The next step was to install the targets at locations predicted during the flight planning stage (see Section 3.5.1). Then, their

positions were measured using the R8 GNSS System (Trimble, Sunnyvale, CA, USA)—a high-precision, dual-frequency RTK system. Following the same concept as camera calibration (Section 3.4.1.2), GCPs were marked as circular targets (Figure 3.5.3a). Once the GCPs were established, the flights for image acquisition started. Table 3.5.1 presents the flight conditions. Labels are given to the acquired datasets for further use in this paper.

Table 3.5.1. Information of the data-acquisition sessions.

Characteristic	Flight Date		
	August 2014	August 2014	October 2014
	Dataset A	Dataset B	Dataset C
Weather temperature (°C)	22	26	10
Wind speed (Km/h)	8	19	8
Zone structure	Stockpiles	Cliffs	Cliffs
Approximate flight altitude (m)	80	90	90

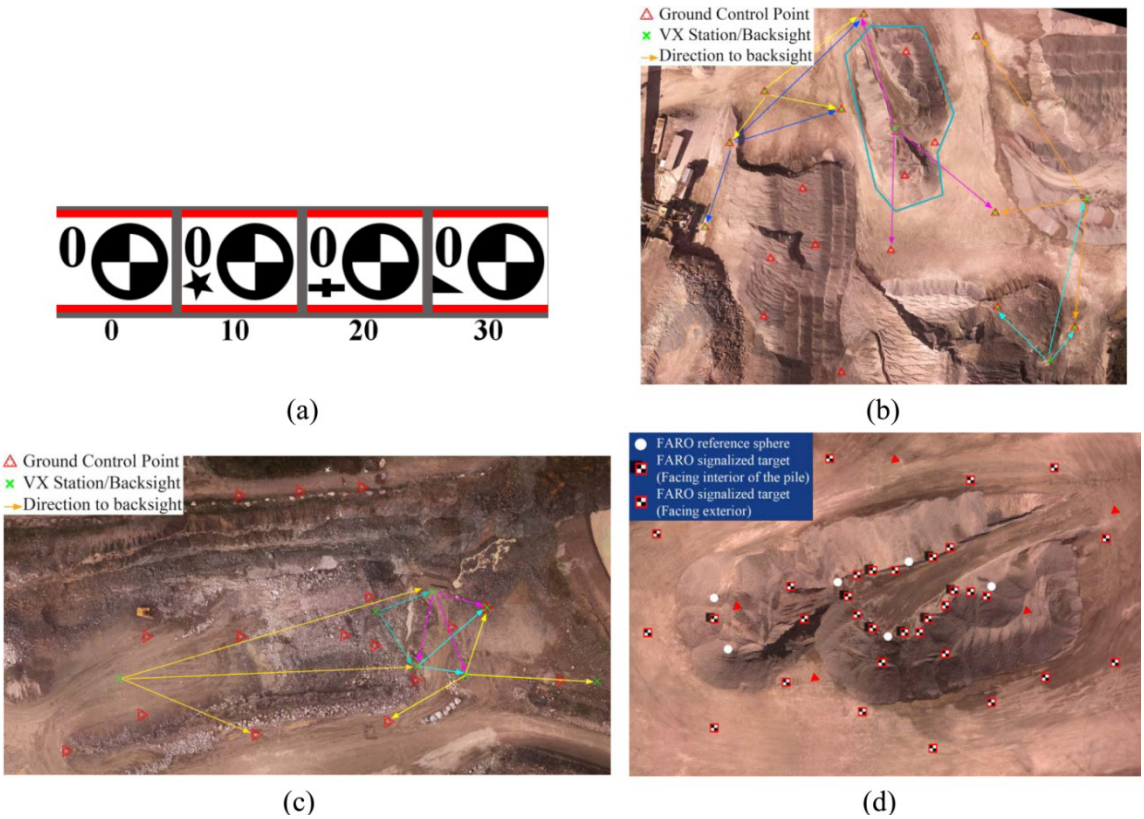


Figure 3.5.3. Surveying plans (**a**) Markers for GCPs. (**b**) Configuration of GCPs and laser-scanner stations for dataset A. (**c**) Configuration of GCPs and laser-scanner stations for dataset C. (**d**) Configuration of targets for FARO scanner over one pile.

Upon termination of image acquisitions, the terrestrial surveying for gathering check data started. The check data included sparse 3D point clouds measured by Trimble VX laser scanner over the whole

mapping area and a dense point cloud measured by a FARO Focus laser scanner (FARO, Lake Mary, FL, USA) over a small pile. Figure 3.5.3b,c show the configuration of the control points, the stations where VX scanner was installed as well as the zone where the dense 3D point cloud was measured by FARO scanner (blue polygon). In order to facilitate the accurate registration of individual FARO scans, several targets with checkerboard pattern and reference spheres were installed at different levels of the pile (Figure 3.5.3d).

3.6 Data Processing Workflow

The main steps of data processing in aerial photogrammetry are illustrated in Figure 3.6.1. The ordinary methods to perform these steps are not discussed here. Instead, the methodology of this study to improve some of these procedures is presented.

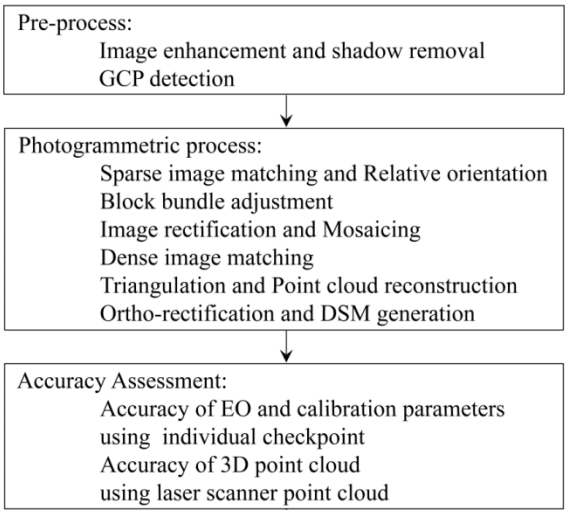


Figure 3.6.1. Photogrammetric workflow to produce topographic data from images.

3.6.1 Image pre-processing

3.6.1.1 Intensity enhancement

As in the system of this study, in most of UAV-PSs, small-format cameras are used because of weight limitations. One of the main problems caused by such cameras is their small ground coverage. This characteristic makes the sequence of images vulnerable to photometric variations, even though the flight time is usually quite short [28]. Noticeable radiometric changes among adjacent images make both sparse matching and pixel-based dense-matching more difficult [29,30]. An example of such situation is given in Figure 3.6.2. These images are from a test dataset not listed in Table 3.5.1. They were taken with a Prosilica GT1920C camera with sensor size of 8.7894 mm × 6.6102 mm and focal length of 16 mm. As shown in Figure 3.6.2a, some images are considerably darker than their neighboring images since a patch of dark clouds had passed through the zone. Some of the images additionally suffer from lack of texture diversity (Figure 3.6.2c). As a result, matching such images becomes difficult.

Therefore, their relative orientation parameters cannot be determined, and the ortho-mosaic cannot be generated either (Figure 3.6.2e).

When thematic applications are required, sophisticated techniques using spectral observations should be used for radiometric correction of images. Otherwise, simple image enhancement methods can be used to reduce the relative photometric variations. In this study, a combination of white balancing and histogram matching is proposed. Starting by the image with proper illumination as the reference image, its dark neighboring image is first white-balanced. Then, the intensity histograms of both images are calculated. If the correlation of the cumulative distribution functions of the histograms is more than 0.5, then no further enhancement is needed. Note that a cumulative histogram represents the intensity rank of each pixel with respect to other pixels regardless of general radiometric changes [31]. If the correlation of two distributions is less than 0.5, then histogram matching is performed to adjust the intensity values in the dark image. With this method, each image is relatively corrected only with respect to the images immediately adjacent to it. Therefore, the trace of dark images is still visible globally. However, this intensity enhancement makes the matching performable, and a correct ortho-mosaic can be generated (Figure 3.6.2f).

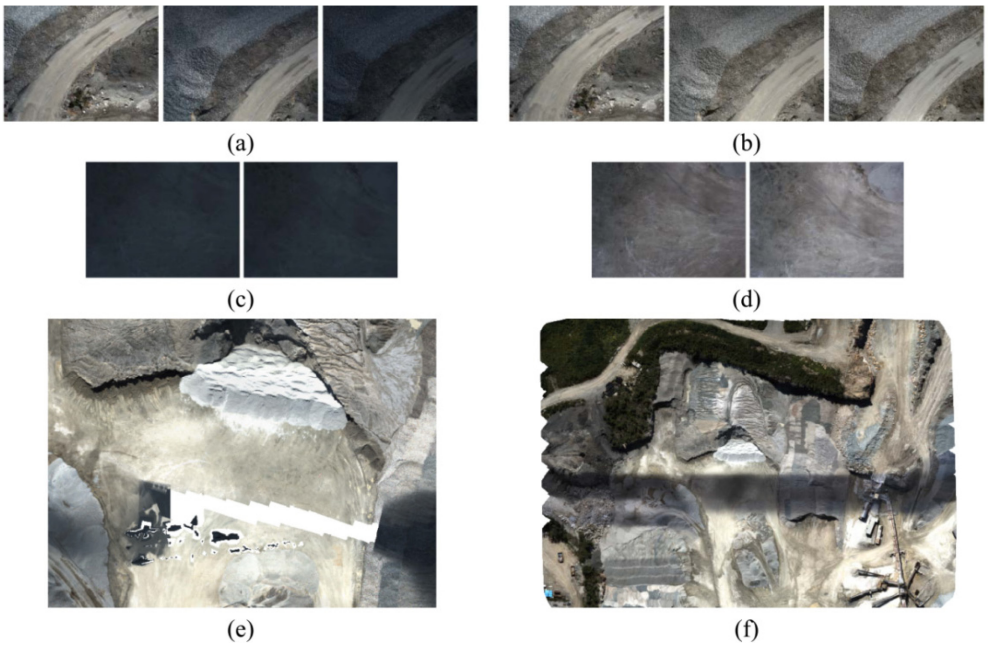


Figure 3.6.2. Intensity enhancement (**a**) A sequence of illuminated and dark images. (**b**) Images of Figure 3.6.2a after correction. (**c**) Two adjacent dark images with low texture diversity. (**d**) Images of Figure 3.6.2c after correction. (**e**) Failure in ortho-mosaic generation. (**f**) Correct mosaic after intensity enhancement.

3.6.1.2 Shadow removal

Another radiometric effect on aerial images is caused by shadows. In thematic applications such as atmospheric correction and classification, shadow regions lead to inevitable errors [32]. Shadows can

also cause spatial errors in 3D modeling. This usually happens when the sun direction changes slightly during the flight and the shadow edges move as a consequence [33]. In our experiments, this problem was observed in dataset B (Table 3.5.1). In order to investigate the effects of shadows on the quality of 3D modeling, a simple technique is proposed to detect and remove the shadow regions from single images. The summary of this technique is presented in Figure 3.6.3. In Section 3.8.3, the effects of shadow removal on both the photometric appearance and the accuracy of the 3D point clouds are analyzed.

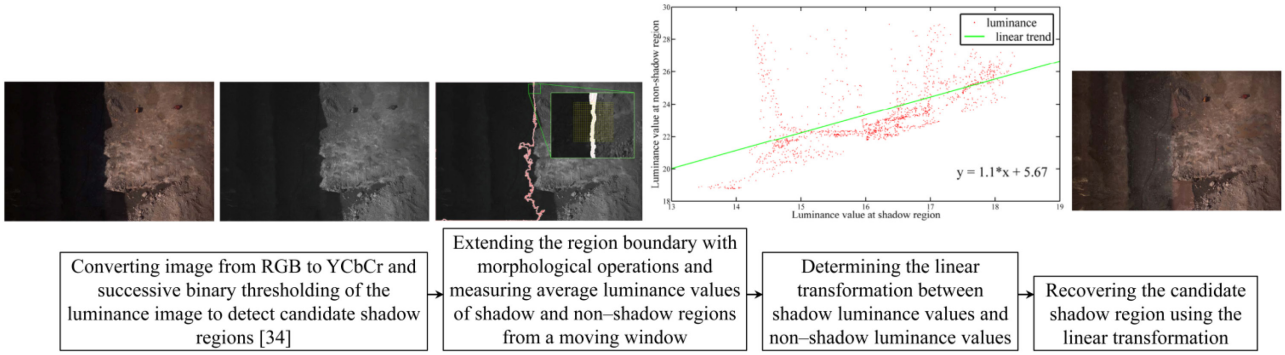


Figure 3.6.3. Workflow for automatic shadow detection and removal.

3.6.1.3 GCP detection

Once the image intensities are enhanced, ground control points can be detected. Automatic detection of GCPs in images is important from two aspects. Firstly, detecting GCPs manually in large sets of UAV images is a cumbersome task. Secondly, the accuracy of target detection directly affects the accuracy of georeferencing and calibration. Therefore, more attention has recently been paid to this process [35,10]. The method applied to position GCPs is presented in Figure 3.6.4. It is mainly based on localization of GCPs using direct EO parameters, color-thresholding and ellipse detection as described in Section 3.4.1.2. Although this process is automatic, the results should manually be verified to remove incorrect detections.

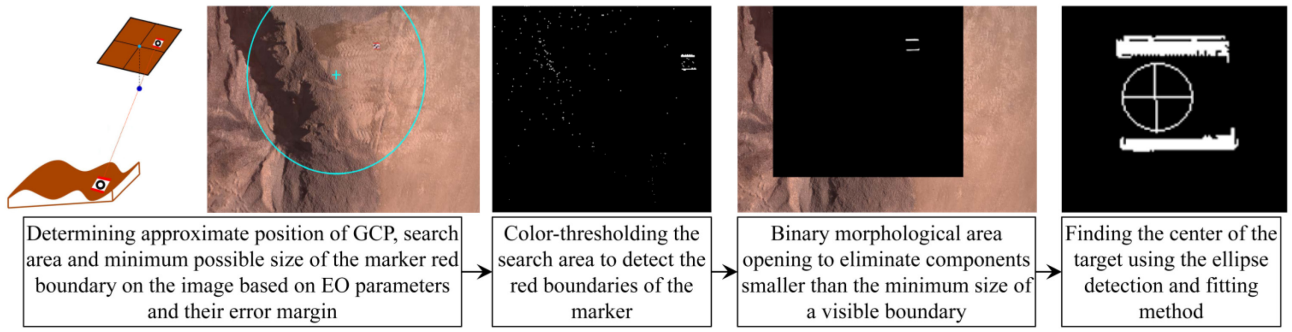


Figure 3.6.4. Workflow for automatic detection of GCPs.

3.6.2 Photogrammetric processing

Photogrammetric processes are applied to aerial images in order to generate various types of topographic products such as 3D point cloud, geo-referenced mosaic and digital surface model (DSM). In this study,

the main photogrammetric processes are performed using Pix4D software [36]. This software has recently been popular among researchers and commercial users for UAV-photogrammetry [33]. Several experiments were designed using this software in order to evaluate the performance of the developed UAV-PS (Section 3.7.2).

Furthermore, a BBA strategy is suggested for on-the-job self-calibration. This strategy is able to control the adverse effects of noisy aerial observations as well as correlation of IO and EO parameters on the accuracy of self-calibration. The main solution of this strategy is to transform the intrinsic camera calibration parameters from unknowns to pseudo-observations. That is, they should be considered as semi-unknowns with known weights. The experimental way to determine these weights is presented in Section 3.7.1, and the results are discussed in Section 3.8.6. In the following paragraphs, the principals of self-calibration with pseudo-observations are presented. Since BBA and least-squares adjustment are well-studied topics, the details are avoided here. Readers are referred to [16,37,38] for more theoretical information.

The mathematical model of bundle adjustment with additional parameters can be presented as in Equation (11), where the observation equations (F) are based on co-linearity condition. These equations are functions of measurements (L), unknowns (Y) and pseudo-observations (X). The measurements are image coordinates of tie points. The unknowns include ground coordinates of tie points and EO parameters of images. The pseudo-observations include intrinsic calibration parameters—both IO parameters and distortion terms:

$$F(X,Y,L)=0 \tag{11}$$

The linear form of Equation (11) is obtained using a Taylor series first-order approximation:

$$W + A\delta K + B\delta L = 0 \tag{12}$$

where K is a concatenated vector by X and Y, W is the miss-closure matrix, and A and B are the matrices of first-order partial derivatives of F with respect to K and L, respectively. Assuming that P_L is the weight matrix of the measurements, P_X is the weight matrix of the pseudo-observations, and D is the matrix of datum constraints, then the least-squares solution for δK can be obtained as in the following equation [38, p. 264, Wells and Krakiwsky, p. 146]:

$$\begin{gathered}\Sigma_K = (P_X + A^T(BP_L^{-1}B^T)^{-1}A + D^TD)^{-1} - D^T(DD^TDD^T)^{-1}D \\ \delta\hat{K} = -\Sigma_K(A^T(BP_L^{-1}B^T)^{-1}W)\end{gathered} \tag{13}$$

The vector of residuals, δL, is also estimated as follows:

$$\delta\hat{\hat{L}} = -P_L^{-1}B^T(BP_L^{-1}B^T)^{-1}(A\delta K + W) \tag{14}$$

These partial solutions, $\delta\hat{K}$ and $\delta\hat{L}$, are successively added to the initial estimations of the unknowns and values of the measurements and pseudo-observations until reaching convergence.

The role of the weight matrix P_X in Equation (13) is to control the changing range of pseudo-observations.

3.7 Experiments

In this section, the experiments which were performed to assess different aspects of the developed system are presented. The results obtained from these experiments are, then, discussed in Section 3.8.

3.7.1 Laboratory experiments

3.7.1.1 Calibration

The camera was calibrated several times during a period of few months before starting the data acquisition. The final parameters were obtained as the average of the parameters from these tests. The stability of each calibration parameter was also determined as its variance at these tests. In the BBA strategy of Section 3.6.2, these variances were used as the weights of pseudo-observations.

In order to verify the accuracy of offline calibration parameters obtained from these test, 10 check images were captured. In these images, the targets were detected using the method of Section 3.4.1.2. Some of them were reserved as checkpoints, and others were served as control points. Using the control points and the calibration parameters, the EO parameters of the images were determined via space resection. Then, the 3D object-space coordinates of the checkpoints were back-projected to the images. The difference between the back-projected position of a checkpoint and its actual position on an image is called the residual. The residuals show how accurate the calibration parameters are modeled. The results obtained from this test are presented in Section 3.8.1.

In order to analyze the efficiency of automatic target detection, compared with manual target detection, similar calibration and assessment tests were performed using the targets that were detected manually. Positions of the manual targets were different from those of the automatic targets with an average of 1.3 pixels and maximum of 2.4 pixels. The results obtained from this experiment are also presented in Section 3.8.1.

3.7.2 Time synchronization

In order to verify how precisely the camera exposures could be tagged via INS messages, a simple experiment was performed. For each image, the INS log file was searched, and the INS message whose GPS time was exactly equal to the GPS timestamp of the image was detected. If the system time tagged to that INS message were adequately close to the system time tagged to the image, then it could be concluded that the GPS timestamp tagged to the image was accurate too. The results of this test, performed on more than 2500 images, are discussed in Section 3.8.2.

3.7.3 Photogrammetric tests

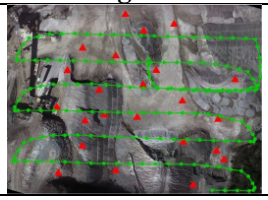
Initially, the images were processed using Pix4D. The 3D triangulated mesh objects for dataset A, dataset B after shadow removal and dataset C are presented in Figure 3.7.1. To evaluate the accuracy of dense reconstruction, cloud-cloud comparison was performed between the image point clouds and

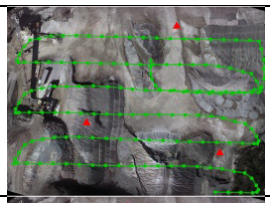
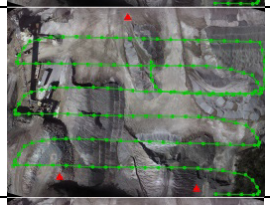
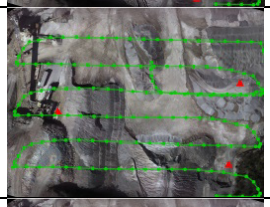
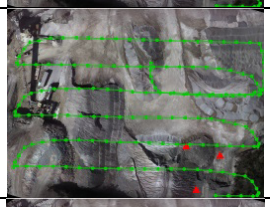
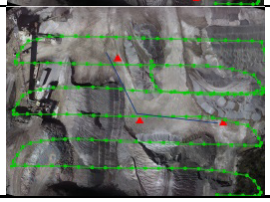
terrestrial laser-scanner point clouds. To this end, the CloudCompare open source software was applied [39]. For each point in the laser cloud, the closest point in the image cloud was found, and the distance between them was calculated. Then, the distances were analyzed to measure the spatial accuracy of image point clouds. The results of these analyses are presented in Section 3.8.3. Once the accuracy of individual point clouds was assessed, they were used to produce other topographic data such as slope maps. Using the DSMs of two different dates, volumetric changes within the site were measured as well. The results are discussed in Section 3.8.7.

Figure 3.7.1. Triangulated mesh for (**a**) Dataset A. (**b**) Dataset B after shadow removal; (**c**) Dataset C.

The second series of the experiments were performed to determine how the number and spatial distribution of GCPs affect the accuracy of 3D modeling. Traditionally, it is known that having more than enough GCPs with good geometrical configuration improves the accuracy of the results [37]. However, in most of UAV-mapping applications, only a minimum number of GCPs can be established. Therefore, it is important to have an *a priori* knowledge of how the final accuracy of 3D modeling would be affected by the GCPs. To this end, several experimental tests were designed. These tests are described through Table 3.7.1. In each test, the initial photogrammetric processing was performed using Pix4D, which included tie point generation, block bundle adjustment and self-calibration. Then, the accuracy of the results was evaluated against checkpoints. The checkpoints in these experiments (Figure 3.7.2) were either the GCPs not used as control points and/or some of the individual laser-scanner points that were transformed to checkpoints. To do this, the ground coordinates of each laser point were back-projected to the images via the accurate indirect EO parameters. Then, SURF feature descriptors were calculated over the projected pixels in all the images [40]. If such pixels represented salient features, and if the distances between their descriptors were smaller than a threshold, then that laser point was considered as a checkpoint. In addition, all the checkpoints were manually verified to avoid errors.

Table 3.7.1. Description of experimental tests for verifying the effect of number/distribution of GCPs.

Test Label	**Figure**	**Descriptions**	
GCPTest 1		Number of GCPs	22
		Distribution	Covering the whole imaging zone

GCPTest 2		Number of GCPs Visibility * Distribution	3 9, 12 and 21 images Evenly distributed over the imaging zone
GCPTest 3		Number of GCPs Visibility Distribution	3 4–6 images Well distributed over the imaging zone
GCPTest 4		Number of GCPs Visibility Distribution	3 19, 20 and 22 images Positioned near the ends of flight strips
GCPTest 5		Number of GCPs Visibility Distribution	3 5, 12 and 21 images Established at the flight home due to inaccessibility to the rest of the imaging zone
GCPTest 6		Number of GCPs Visibility Distribution	3 15, 17 and 22 images Established along a hypothetical road due to inaccessibility to other areas

* Number of images, in which every GCP is visible.

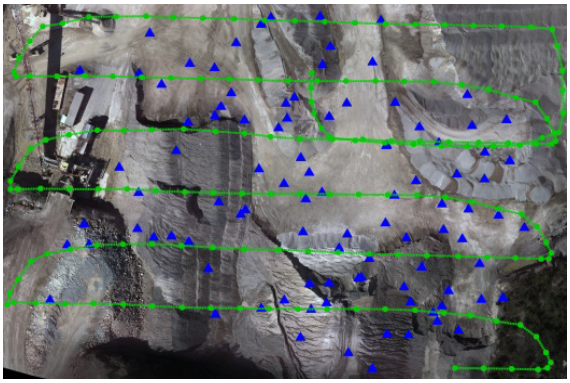


Figure 3.7.2. Flight trajectory and distribution of checkpoints in dataset A.

Moreover, we were interested in assessing the results of these tests not only with GCPs that were measured by RTK GPS, but also with GCPs whose positions were measured by other techniques. Since the GCPs were originally measured only by accurate RTK GPS, it was decided to simulate the measurements for other techniques. To this end, three reference points were established outdoor, and

their exact coordinates with RTK GPS system were measured. Then, other GPS devices were used to re-measure their coordinates. Using the observations made by each device, the positioning errors of GCPs were simulated. First, a Garmin GLO-GPS was used, and more than 10,000 observations were recorded over the reference points. This device is WAAS-enabled and receives position information from both GPS and GLONASS satellites. The root mean square (RMS) positioning error for this device was 2.40 m horizontally and 6.04 m vertically. Similarly, a series of 2000 observations were made with a SXBlueII GPS. This device is also WAAS-enabled and performs additional code-phase measurements and multi-path error reduction. The RMS positioning error with this device was 0.65 m horizontally and 0.69 m vertically. The results obtained in different tests using these types of GCPs are presented in Section 3.8.4.

To analyze the effect of imaging configuration in absence/presence of GCPs on the accuracy of 3D modeling, a sequence of nine images was considered (Figure 3.7.3a). Ground control points A, B and C were visible in images 1–4, while point D was visible in images 5–6. The following situations were, then, designed and tested.

i. OverlapTest 1: Each image was overlapped with at least three connected images. For example, image 5 had common tie points with both images 3 and 4. Such a connection is illustrated via the connectivity matrix in Figure 3.7.3b.

ii. OvelapTest 2: The situation was the same as OverlapTest 1. However, image 5 did not have any common tie points with both images 3 and 4; *i.e.*, it was only overlapped with image 4. The connectivity matrix of Figure 3.7.3c shows this situation.

The objective of these tests was to find out whether the lack of overlap in OverlapTest 2 could cause problems in BBA, and how important the role of GCPs was to solve those problems. The results obtained from each test and the issues involved with each situation are assessed in Section 3.8.5.

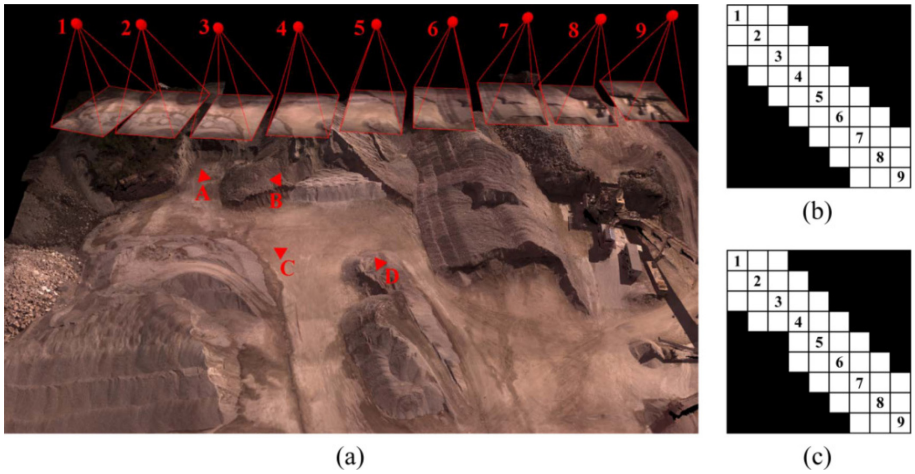


Figure 3.7.3. (**a**) Experiment to assess the effect of imaging configuration. (**b**) Connectivity matrix in OverlapTest 1. (**c**) Connectivity matrix in OvelapTest 2.

The final series of photogrammetric experiments were performed to analyze the effect of on-the-job self-calibration on the accuracy of IO and EO parameters. Also, the proposed self-calibration strategy of Section 3.6.2 was evaluated, and the results were compared with those of traditional self-calibration. To this end, the following situations were considered, and correlation analysis was performed at each situation in order to determine the dependency of IO parameters to EO parameters.

i. CalibTest 1: Offline camera calibration was performed using a well-configured imaging network (Figure 3.7.4a).

ii. CalibTest 2: On-the-job calibration was performed using typical aerial images, which were all acquired from almost the same altitude (Figure 3.7.4b).

iii. CalibTest 3: On-the-job calibration was performed using aerial images, which were acquired from varying altitudes (Figure 3.7.4c).

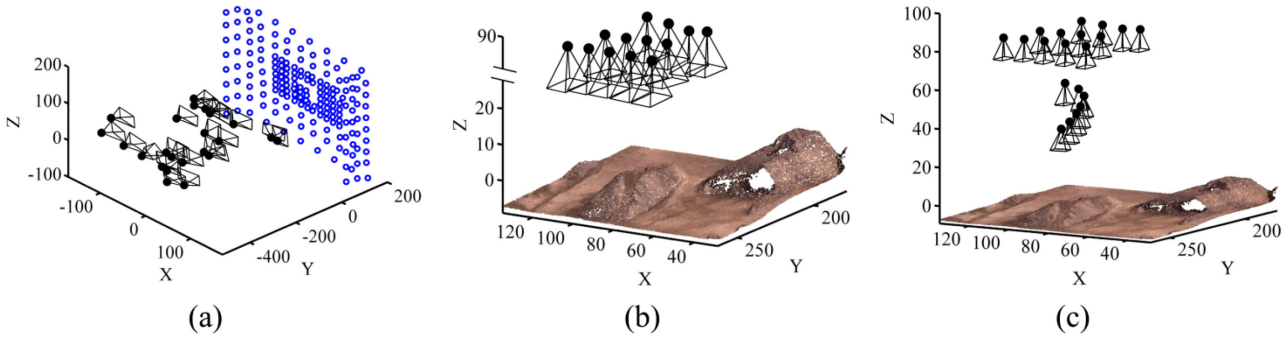


Figure 3.7.4. Self-calibration experiments (**a**) CalibTest 1. (**b**) CalibTest 2. (**c**) CalibTest 3.

3.8 Results and Discussion

3.8.1 Calibration results

Figure 3.8.1a shows the residual vectors for the checkpoints after camera calibration. The mean and standard deviation (StD) of the residuals on the checkpoints at *x*- and *y*-directions are 0.32 ± 0.18 pixel and 0.20 ± 0.16 pixel, respectively. Figure 3.8.1b presents the residual vectors based on the manual target detection. Notice that the targets on check images were detected automatically and, only, the targets used for calibration were measured manually. As a result, the automatic target detection improves the accuracy of calibration 81.25% in comparison with the noisy observations based on manual detection even though the noise level does not exceed 2.4 pixels.

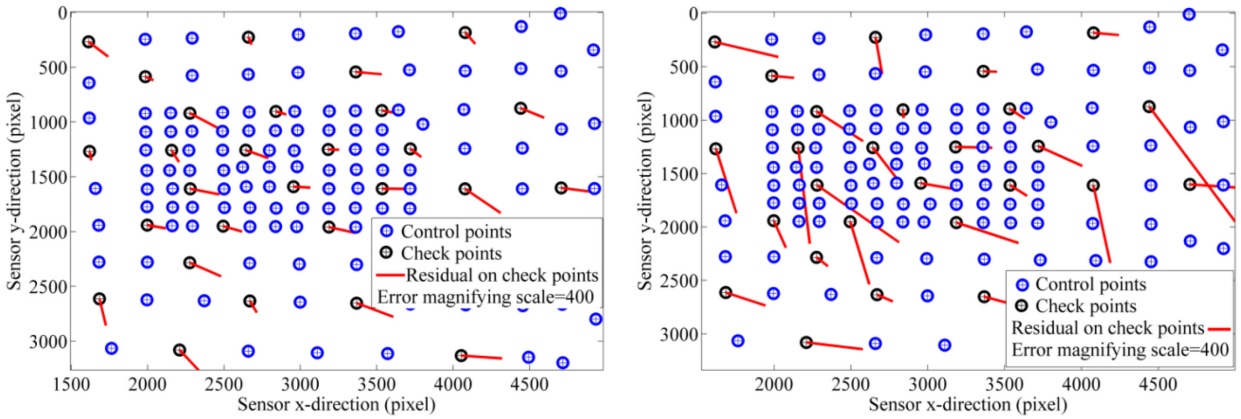


Figure 3.8.1. Calibration results (**a**) Residuals on checkpoints based on automatic target detection. (**b**) Residuals on checkpoints based on manual target detection.

3.8.2 Precision of time-synchronization

Figure 3.8.2 demonstrates an example of the results obtained from the time-synchronization test. The *x*-axis shows the image number. The *y*-axis shows ΔT, which is the absolute difference between the system time tagged to each image and the system time tagged to its corresponding INS message. It is, indeed, the difference between the real exposure-end time of an image and the GPS timestamp tagged to it. As it can be noticed, these differences are random; however, they do rarely exceed 20 ms. This is due to the fact the INS frequency is 50 Hz (1 message per 20 ms). The main reason why this difference (ΔT) is random is that the camera exposures do not start on very exactly fixed intervals—e.g., every 500 ms. Instead, there is a few milliseconds of random delay/advance from the defined interval—e.g., 502 ms. In average, it can be concluded that the GPS timestamp tagged to any image is approximately 11 ± 7 ms delayed/advanced from the exact time of the exposure.

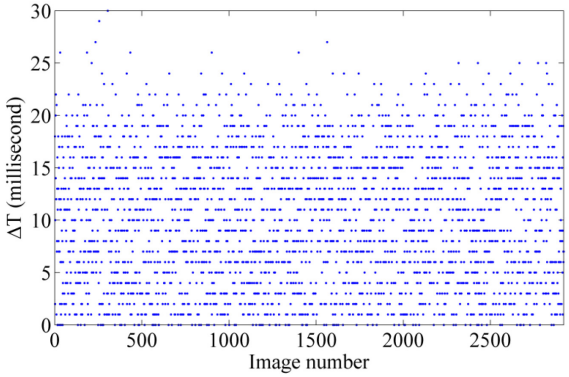


Figure 3.8.2. Results of the time-synchronization test.

3.8.3 Accuracy of 3D point Clouds

As mentioned in Section 3.7.2, image point clouds were compared with terrestrial laser-scanner point clouds. Figure 3.8.3a,b illustrate the histograms of horizontal and vertical distances between the point cloud of dataset A and that of the laser scanner (see Table 3.8.1 as well). The vertical accuracy of this point cloud is 1.03 cm, and its horizontal accuracy is 1.58 cm.

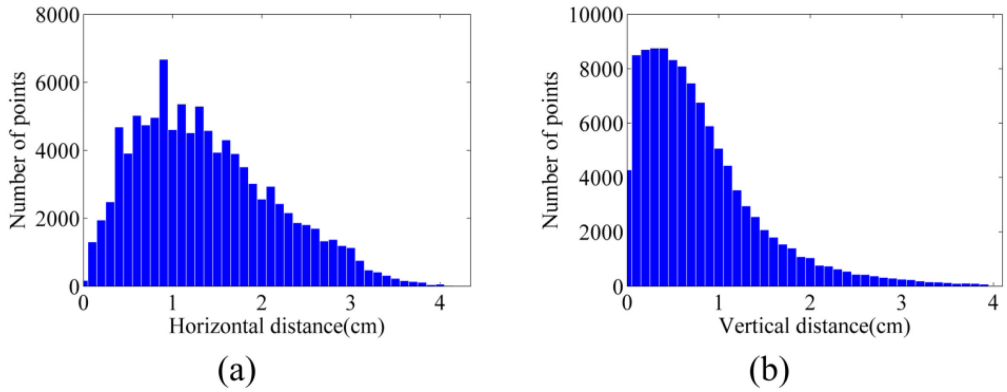


Figure 3.8.3. Histograms of distances between the point clouds from dataset A and laser scanner (**a**) Horizontal distances. (**b**) Vertical distances.

Table 3.8.1. Summary of distances between the image-based point clouds and laser-scanner ones.

	Horizontal Distance (cm)			**Vertical Distance (cm)**		
Dataset	**Mean**	**RMS**	**StD**	**Mean**	**RMS**	**StD**
A	1.38	1.58	0.77	0.80	1.03	0.66
B before shadow removal	1.79	2.03	0.96	1.41	1.72	0.99
B after shadow removal	1.62	1.82	0.83	1.32	1.62	0.95
C	1.88	2.07	0.84	1.63	2.02	1.18

As another test, the dense point cloud measured by FARO laser scanner was transformed to a raster DSM. Then, the absolute difference of the image-based DSM from the laser DSM was calculated. Figure 3.8.4 presents the vertical difference between the image-based DSM and laser DSM. In more than 78% of the zone, this vertical difference is less than 1 cm.

For dataset B, as mentioned in Section 3.6.1.2, shadow regions were removed from the images. Figure 3.8.5a,b show the geo-referenced mosaics of the site before and after shadow removal, respectively. As it can be seen, the results are visually improved. The shadow-free mosaic can be used in thematic applications where shadow effects cause errors.

The point clouds obtained before and after shadow removal were also evaluated using the laser-scanner data. The results are presented in Table 3.8.1 and

Figure 3.8.6. According to the results, the vertical accuracy is improved from 1.72 cm to 1.62 cm with shadow removal. Therefore, no noticeable improvement can be observed via this test. This is principally

due to the fact that the terrestrial laser-scanner points over the shadow region were not dense enough. However, when the DSMs before and after shadow removal were compared, the differences could be observed more clearly. As shown in Figure 3.8.7, large vertical differences, as large as 8 cm, can be observed in edges of the cliffs. These are the zones where more than one shadow was casted on the objects—the shadow from two higher rows of the rocks. It is believed that this type of shadow causes errors in the dense matching and 3D reconstruction process.

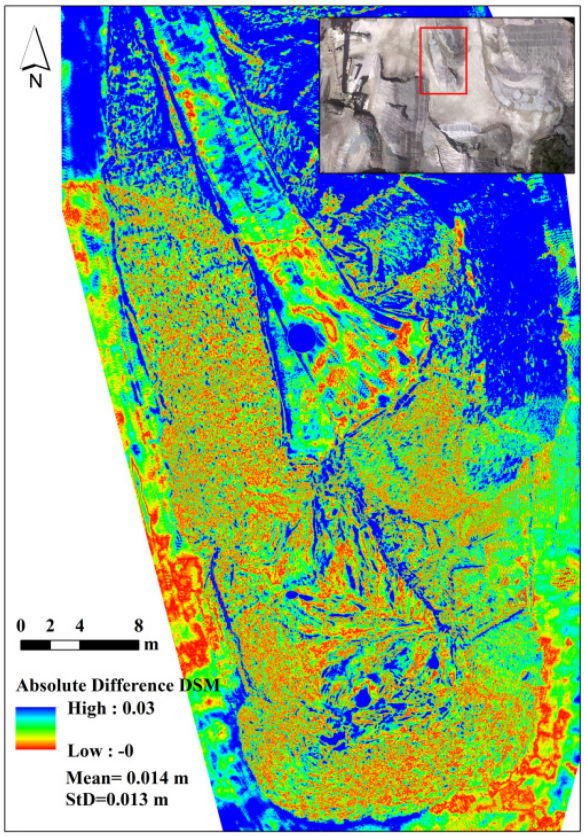
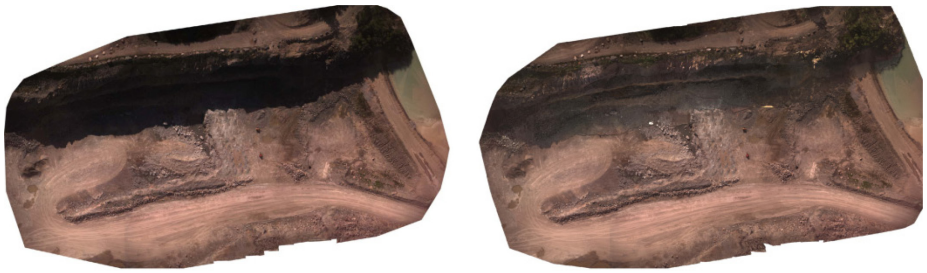


Figure 3.8.4. Absolute difference between the DSM from FARO laser scanner and that of dataset A.



(a) (b)

Figure 3.8.5. Image mosaics of dataset B (**a**) Before shadow removal. (**b**) After shadow removal.

Finally, the point cloud of dataset C was evaluated against the terrestrial laser point cloud. The results are presented in Table 3.8.1 and Figure 3.8.8. Generally, the accuracy of the point cloud from dataset A is higher than both dataset B and dataset C. This could be explained by the structure of the stockpiles in dataset A, where less occlusion happens in aerial images. For datasets B and C, the vertical and layered structure of the cliffs causes more errors.

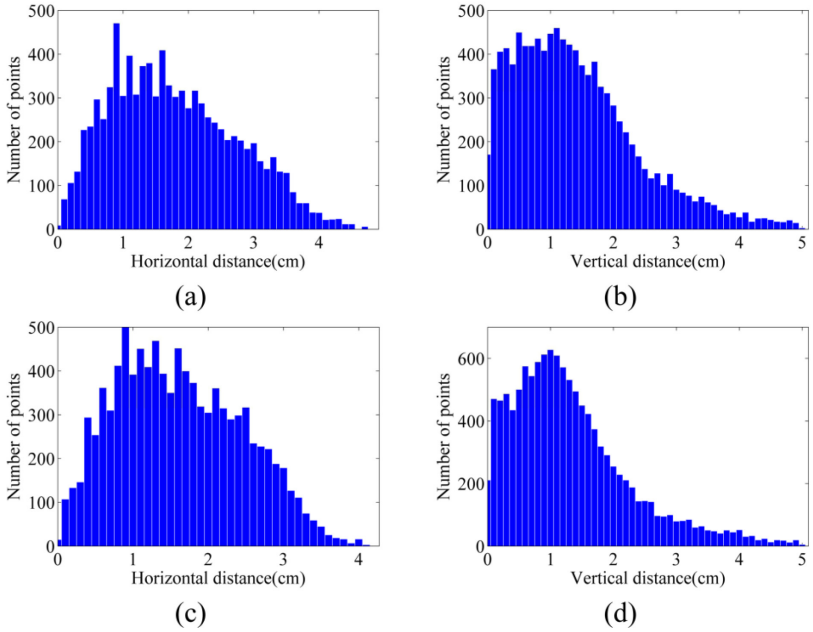


Figure 3.8.6. Histograms of distances between the point clouds from dataset B and laser scanner (**a**) Horizontal distances before shadow removal. (**b**) Vertical distances before shadow removal. (**c**) Horizontal distances after shadow removal. (**d**) Vertical distances after shadow removal.

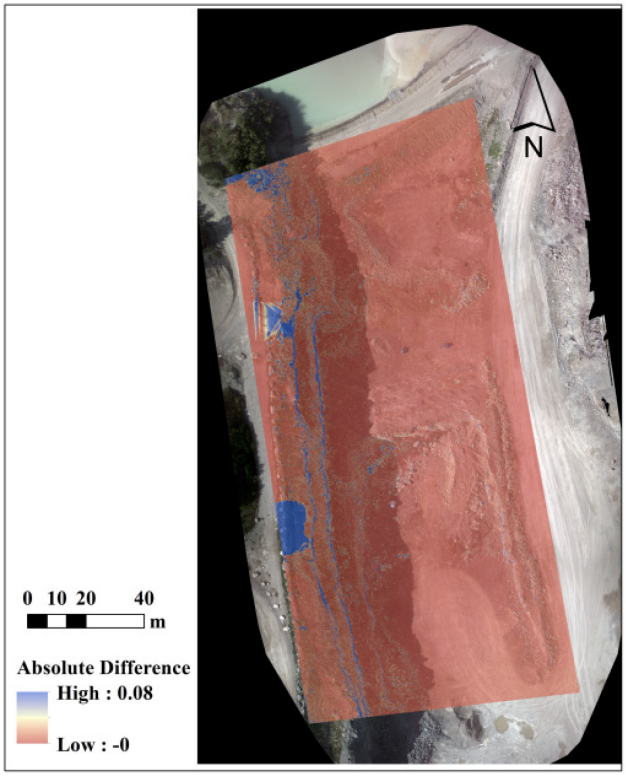


Figure 3.8.7. Absolute difference between the DSMs of dataset B before and after shadow removal.

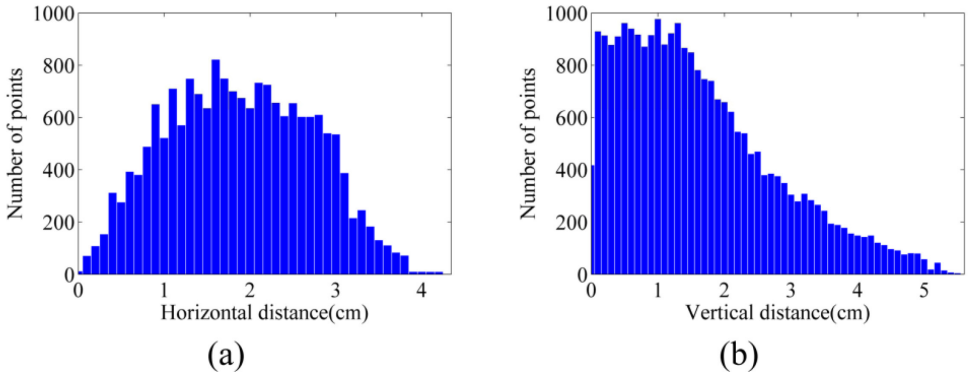


Figure 3.8.8. Histograms of distances between the point clouds from dataset C and laser scanner (**a**) Horizontal distances. (**b**) Vertical distances.

3.8.4 Effects of ground control points

Firstly, the results obtained via direct georeferencing are discussed. In fact, DG may be interpreted in two senses. In the first sense, direct EO parameters of images from the navigation data are directly used for 3D modeling without any photogrammetric refinement applied to them. In this case, offline intrinsic

camera calibration parameters should be used as no on-the-job self-calibration is performed. This strategy is mostly used for rapid mosaicking and ortho-photo generation. The results from this test for dataset A are shown as DG type 1 in Table 3.8.2. In the second sense, direct navigation data are used as inputs into initial photogrammetric processing and the EO parameters are slightly refined within a free-network adjustment. Then, these refined EO parameters are used for 3D modeling. The results from this test are shown as DG type 2 in Table 3.8.2. Notice that in this table and the following ones, the mean error represents the average of absolute errors—not the signed ones. As it can be seen, the horizontal and vertical accuracy is improved 31% and 73%, respectively, after applying the initial processing. The main reason for this improvement is that all the relative-orientation errors between images are corrected within the initial processing. This can be observed from the fact that standard deviations of errors decrease considerably after initial processing.

Table 3.8.2. Accuracy of direct georeferencing on checkpoints.

Experiment	X-Direction Error (m)			Y-Direction Error (m)			Vertical Error (m)		
	Mean	RMS	StD	Mean	RMS	StD	Mean	RMS	StD
DG type 1	1.146	1.406	0.837	2.478	3.088	1.892	10.440	11.670	5.359
DG type 2	1.938	1.943	0.137	1.162	1.166	0.090	3.159	3.169	0.268

The results obtained from the experimental tests with GCPs (Table 3.7.1, Section 3.7.2) are presented in Table 3.8.3. For each checkpoint, the EO and calibration parameters of images—after indirect geo-referencing at each experimental test—were used to determine its ground coordinates via intersection. Then, the error on the checkpoint was measured as the difference between its ground-truth 3D coordinates and the calculated coordinates.

In order to provide a better understanding of the way each configuration or device affects the results, the relative changes of accuracy are represented in Table 3.8.4. These change rates are calculated as the percentage of RMS improvement with regard to the lowest accuracy. Therefore, the improvement rate of 0.0% shows the reference value used for change-percentage measurement.

As the results show, in order to reach the highest accuracy, it is recommended to provide a large number of well-distributed GCPs (as in GCPTest 1). However, if this is not possible, then the best solution is to install the GCPs at different sides of the imaging zone, where they can also be visible in as many images as possible (as in GCPTest 2). To ensure this condition, the best practice is to install them near the ends of the flight strips so that they are visible in several images from two adjacent strips (as in GCPTest 4). Typically, it is preferred to install GCPs at places with height variation. However, the results from GCPTest 6, where the control points are almost at the same elevation, are much more accurate than those of GCPTest 3, where GCPs have high height variation but low visibility. Finally, the least accurate results are obtained from GCPTest 5, where the GCPs are positioned at the flight home. Therefore, this solution should be avoided unless there is no other possibility. Besides, in this situation, it should be ensured that the GCPs can be commonly visible in at least three images. In order to plan any of these situations, the flight-planning software (Section 3.5.1) can be used.

Table 3.8.3. Horizontal and vertical accuracy on checkpoints based on different GCP experiments.

Error	Experiment	Trimble R8			SXBlue			Garmin GLO		
		Mean	RMS	StD	Mean	RMS	StD	Mean	RMS	StD
Horizontal Error (cm)	GCPTest 1	0.2	0.4	0.3	61.9	61.9	3.0	180.0	180.7	12.4
	GCPTest 2	0.3	0.4	0.6	68.0	69.0	1.7	158.2	160.8	19.6
	GCPTest 3	0.8	0.9	1.2	73.9	74.1	4.6	216.4	216.6	9.0
	GCPTest 4	0.3	0.4	0.7	63.8	62.9	2.3	160.2	165.3	30.2
	GCPTest 5	0.6	0.8	0.9	74.7	76.3	14.8	227.0	228.0	20.5
	GCPTest 6	0.3	0.5	0.6	72.6	72.8	5.1	189.3	193.5	31.4
Vertical Error (cm)	GCPTest 1	1.2	1.7	1.2	13.8	15.5	7.0	412.9	413.0	10.5
	GCPTest 2	1.6	2.0	1.2	41.1	49.7	28.3	355.5	355.8	12.4
	GCPTest 3	4.1	4.3	1.4	73.6	73.6	2.8	434.9	436.0	32.5
	GCPTest 4	1.4	2.0	1.4	43.2	48.5	22.1	432.1	433.5	35.2
	GCPTest 5	2.4	3.0	1.8	121.6	147.1	83.6	431.2	446.1	115.6
	GCPTest 6	1.4	1.9	1.4	80.2	97.4	55.7	432.1	433.5	35.2

Table 3.8.4. Improvement rate of accuracy on checkpoints based on different GCP experiments.

Device	Horizontal-Accuracy Percentage Change						Vertical-Accuracy Percentage Change					
	Experiment						Experiment					
	1 *	2	3	4	5	6	1	2	3	4	5	6
R8 RTK	99.8	99.8	99.6	99.8	99.6	99.8	99.6	99.6	99.0	99.6	99.3	99.6
SXBlue	72.9	69.7	67.5	72.4	66.5	68.1	96.5	88.9	83.5	89.1	67.0	78.2
Garmin GLO	20.7	29.5	5.0	27.5	0.0	15.1	7.4	20.2	2.3	2.8	0.0	2.8

* Reads as GCPTest 1.

3.8.5 Effects of imaging configuration

The above-mentioned experiments prove that careful application of minimum GCPs can also yield a high level of modeling accuracy. However, such accuracy level is only achievable if images provide a stable imaging and network configuration. The importance of this fact is analyzed based on the overlap tests described in Section 3.7.2.

When performing BBA, the coordinate datum requires seven defined elements to compensate its rank deficiencies, namely scale, position and rotation. These defined elements can be provided with either minimum constraints in controlled networks or inner constraints in free networks. When enough overlap exists among images, both free and controlled network adjustments can be performed correctly without facing any additional rank deficiencies. Figure 3.8.9a shows the orientations of cameras and ground coordinates of tie points calculated correctly in a free network based on OverlapTest 1.

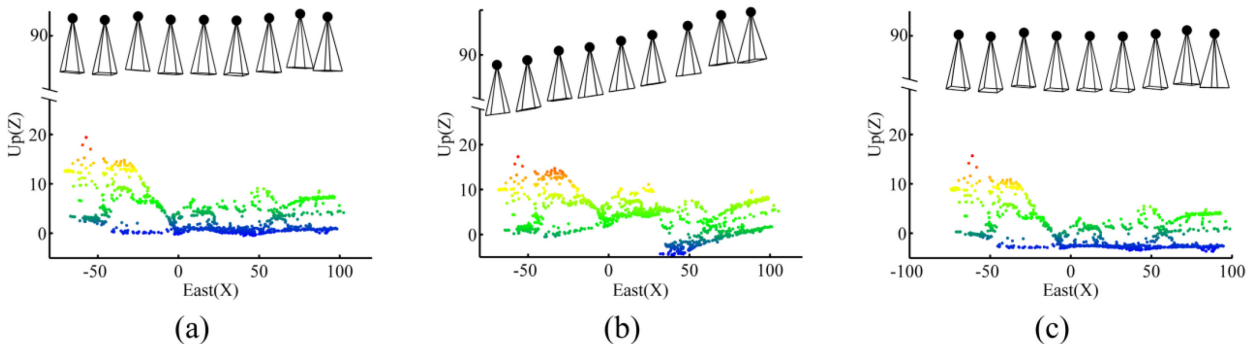


Figure 3.8.9. Effect of imaging configuration (**a**) Inner-constrained adjustment based on OverlapTest 1. (**b**) Inner-constrained adjustment based on OverlapTest 2. (**c**) Controlled adjustment based on OverlapTest 2 using four GCPs.

However, only one image not having enough overlap with its adjacent ones can disturb this ideal configuration. As in OverlapTest 2, image 5 does not have any common tie point with image 3. Therefore, there is no tie point to make a connection between one part of the network including images 1–4 and the other part of the network including images 5–9. Notice that this disconnection happens even though image 5 and image 4 have common tie points. As a result, the coordinate datum faces eight rank deficiencies—one additional scale deficiency. In order to resolve this, one more constraint is required. If no ground-truth measurement is available, then the solution is to assign an arbitrary scale factor to one of the unknowns. Figure 3.8.9b illustrates the results by assigning a scale factor based on the DG data to one of the tie points between images 4 and 5. In this situation, although the BBA can be solved, a wrong scale change is introduced between the two parts of the network. As a result, this solution must be avoided unless the DG data are very accurate. The practical solution to this problem is to add the ground observations of control points to the adjustment. In this example, control point D provides the additional scale constraint required to solve the 8th rank deficiency of the datum (Figure 3.8.9c). It can be concluded, that configurations of both the terrestrial data (GCPs or any other types of ground measurements) and the aerial images decide the final accuracy of 3D modeling.

3.8.6 On-the-job self-calibration results

As the results in Section 3.8.1 show, noisy image observations affect the results of self-calibration to a great extent even if the noise level is very low. Similarly, on-the-job self-calibration of aerial images is affected by the noise in images, which is usually inevitable in UAV imagery. Another factor that affects the accuracy of on-the-job self-calibration is the particular configuration of aerial network. That is the images are acquired from a relatively fixed altitude. In fact, this network configuration reduces the numerical stability of calibration in terms of the increase in the correlation between the unknown parameters. Especially, IO parameters—the principal point offset and focal length (x_p, y_p, f)—become strongly correlated with EO parameters—the position of the camera center (C_x, C_y, C_z). As a result, intrinsic calibration parameters become physically meaningless since they become dependent parameters that change relatively with the changes of EO parameters.

This effect can be practically controlled in close-range photogrammetry by providing various orientations and object depth levels as in CalTest 1. Figure 3.8.10a presents the correlation analysis of the self-calibration based on these images. As it is noticed, this condition results in very low correlation between IO and EO parameters. However, the same analysis for on-the-job self-calibration based on aerial images of CalTest 2 presents very high correlation between IO and EO parameters, specifically between focal length and imaging depth (Figure 3.8.10b).

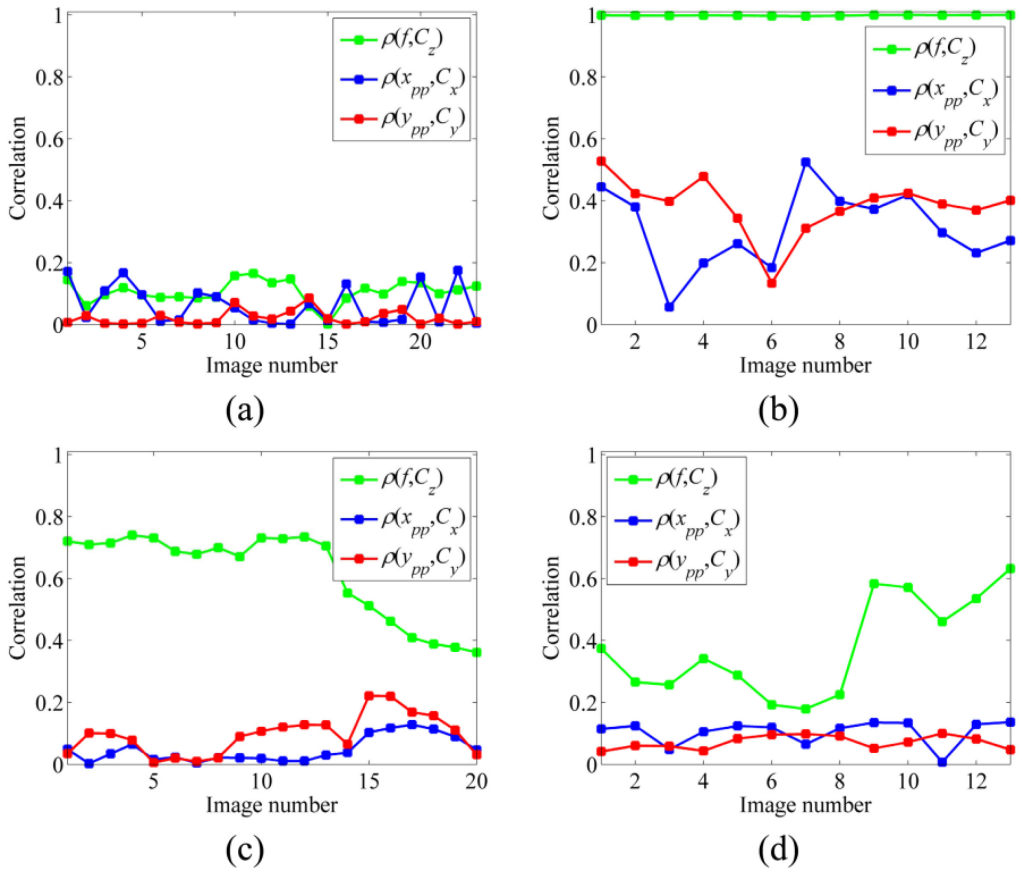
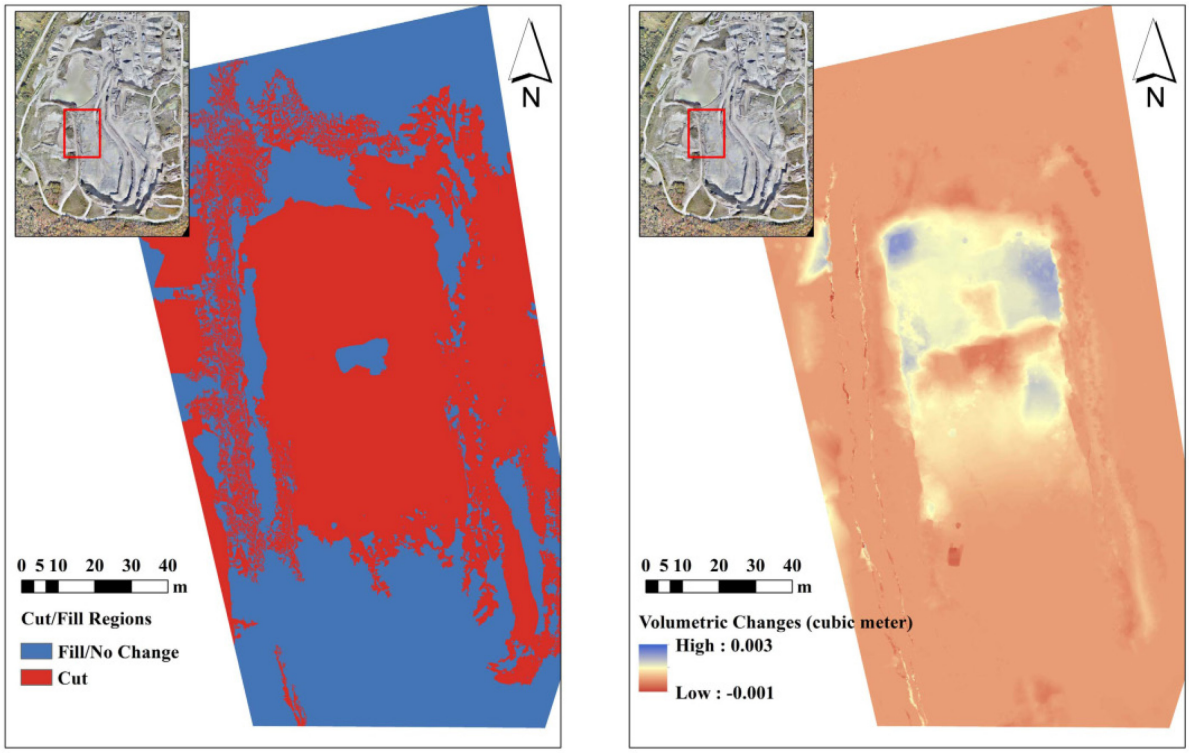


Figure 3.8.10. Correlation analysis in self-calibration based on (**a**) CalTest 1. (**b**) CalTest 2. (**c**) CalTest 3. (**d**) CalTest 2 by applying the proposed BBA strategy.

One of the advantages of UAVs is that they can fly obliquely and at very low altitudes. Therefore, it is possible to provide more orientation variations as in CalTest 3. As a result, the average correlation between focal length and imaging depth can be reduced 38% by this new configuration (Figure 3.8.10c). However, such maneuvers are not possible in all the UAV mapping applications. Therefore, the solution proposed in Section 3.6.2 can be applied to improve the self-calibration. This strategy reduces the correlation between the unknowns without the need to change the network configuration (Figure 3.8.10d). For instance, the average correlation between focal length and imaging depth is reduced 60%.

3.8.7 Application-dependant results

Figure 3.8.11a presents the major cut/fill regions based on dataset B and dataset C that were gathered with an interval of two months. As expected, most places at this zone were excavated. In Figure 3.8.11b, the volumetric change per cell is measured for every cell of the DSM. The volumetric change is measured as the difference of elevation in the before-DSM from the after-DSM which is multiplied by the cell area (1.69 cm^2). Therefore, positive values represent excavation or cut, and negative values represent fill. The vertical accuracies of dataset B (before) and dataset C (after) are 1.32 cm and 1.63 cm, respectively. Therefore, the volumetric change measurement at each cell is performed with accuracy of 3.54 cm^3. Figure 3.8.12 presents the slope map based on dataset A. As it can be seen, very detailed slope information is extractable from such a map, which can be used in various geological applications.



(a) (b)

Figure 3.8.11. (**a**) Cut/fill regions. (**b**) Volumetric change measurement.

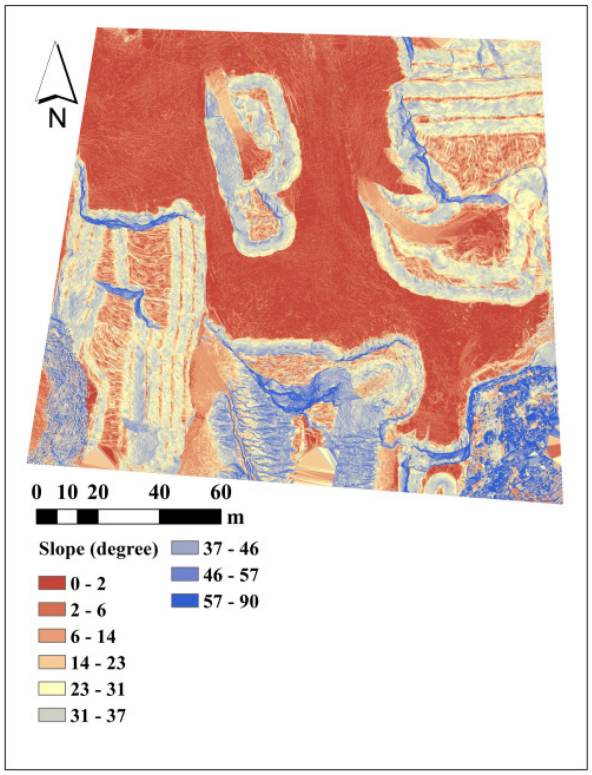


Figure 3.8.12. Classified slope map based on dataset A.

3.9 Conclusions

Various aspects of the development and implementation of a UAV-PS were discussed in this study. These included the camera offline calibration, platform calibration, system integration, flight and fieldwork planning, data acquisition, photogrammetric processing and application in open-pit mine mapping. Based on the experiments, it was concluded that the accuracy of 3D modeling with the system, either in terms of the accuracy of indirect georeferencing or the spatial accuracy of the point clouds, was better than 2 cm.

In addition to general photogrammetric experiments, several tests were performed to analyze the specific issues of UAV-based 3D modeling, and solutions were proposed to address them. It is hoped that the lessons learnt from these experiments give a more clear insight of the capacities of UAV-PSs for the upcoming studies and applications. In brief, the impact of automatic target detection on the accuracy of camera calibration was investigated. It was shown that an improvement of 81% in the accuracy of calibration could be achieved with our target detection technique in comparison with manual target detection. Regarding the system integration, it was validated that the developed software package was capable of synchronizing the navigation and imaging sensors with an approximate delay of 11 ms without requiring any additional mechanism. Moreover, the impacts of high photometric variations among images and shadowed regions on the accuracy of 3D modeling were verified. Besides, the use of a BBA strategy was suggested to improve the accuracy of on-the-job self-calibration by reducing the

correlation of intrinsic camera calibration parameters to other BBA elements such as EO parameters. It was shown that, using this strategy, the correlation of IO and EO parameters could be reduced by 60% in an unsuitable imaging network. This strategy can be used in applications where the accurate, on-the-flight intrinsic calibration parameters are required independently. Furthermore, several experiments were performed to assess the effect of GCPs configuration on modeling accuracy. It was shown that a minimum number of GCPs could provide a high accuracy level if they were distributed evenly over the whole zone and their visibilities in images were maximized. However, under such conditions, the scale consistency of the imaging network needed to be ensured by providing high overlap among images.

Acknowledgments

This study is supported by grants from the: Centre de Géomatique du Québec, Fonds de Recherche Québécois sur la Nature et les Technologies, and Natural Sciences and Engineering Research Council of Canada. The authors would also like to thank Kim Desrosiers from Centre de Géomatique du Québec who took time from his schedule to help us optimize the system-integration software.

Author Contributions

Mozhdeh Shahbazi performed the theoretical developments and planning of this system. Unless otherwise mentioned in the paper, she wrote the software packages for implementing the system, performing the experiments and analyzing the results. Furthermore, she prepared the first draft of this manuscript. The manuscript was revised and modified by the other authors as well Patrick Ménard performed the hardware development. All the authors assisted in data acquisition and fieldworks. Gunho Sohn and Jérôme Théau provided the direction and supervision of this study. They provided their constructive comments and consults to perform this study.

Conflicts of Interest

The authors declare no conflict of interest.

3.10 References

1. Shahbazi, M.; Théau, J.; Ménard, P. Recent applications of unmanned aerial imagery in natural resource management. *Gisci. Remote Sens.* **2014**, *51*, 339–365.
2. Anderson, K.; Gaston, K.J. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Front. Ecol. Environ.* **2013**, *11*, 138–146.
3. Liu, P.; Chen, A.; Huang, Y.; Han, J.; Lai, J.; Kang, S.; Wu, T.; Wen, M.; Tsai, M. A review of rotorcraft unmanned aerial vehicle (UAV) developments and applications in civil engineering. *Smart Struct. Syst.* **2014**, *13*, 1065–1094.

4. Anai, T.; Sasaki, T.; Osaragi, K.; Yamada, M.; Otomo, F.; Otani, H. Automatic exterior orientation procedure for low-cost UAV photogrammetry using video image tracking technique and GPS information. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, doi:10.5194/isprsarchives-XXXIX-B7-469-2012.
5. Bahr, T.; Jin, X.; Lasica, R.; Giessel, D. Image registration of high-resolution UAV data: The new hypare algorithm. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, doi:10.5194/isprsarchives-XL-1-W2-17-2013.
6. Boccardo, P.; Chiabrando, F.; Dutto, F.; Tonolo, F.G.; Lingua, A. UAV deployment exercise for mapping purposes: Evaluation of emergency response applications. *Sensors* **2015**, *15*, 15717–15737.
7. Turner, D.; Lucieer, A; Wallace, L. Direct georeferencing of ultrahigh-resolution UAV imagery. *IEEE Trans. Geosci. Remote* **2014**, *52*, 2738–2745.
8. Xiang, H.; Tian, L. Method for automatic georeferencing aerial remote sensing (RS) images from an unmanned aerial vehicle (UAV) platform. *Biosyst. Eng.* **2011**, *108*, 104–113.
9. Chiang, K.W.; Tsai, M.L.; Chu, C.H. The development of an UAV borne direct georeferenced photogrammetric platform for ground control point free applications. *Sensors* **2012**, *12*, 9161–9180.
10. Turner, D.; Lucieer, A.; Watson, C. An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote Sens.* **2012**, *4*, 1392–1410.
11. Chiang, K.W.; Tsai, M.L.; Naser, E.S.; Habib, A.; Chu, C.H. New calibration method using low cost MEMS IMUs to verify the performance of UAV-borne MMS payloads. *Sensors* **2015**, *15*, 6560–6585.
12. Ai, M.; Hu, Q.; Li, J.; Wang, M.; Yuan, H.; Wang, S. A robust photogrammetric processing method of low-altitude UAV images. *Remote Sens.* **2015**, *7*, 2302–2333.
13. Wu, C.T.; Hsiao, C.Y.; Chen, C.S. An assessment of errors using unconventional photogrammetric measurement technology with UAV photographic images as an example. *J. Appl. Sci. Eng.* **2013**, *16*, 105–116.
14. Remondino, F.; Fraser, C. Digital cameras calibration methods: Considerations and comparisons. In Proceedings of the ISPRS Commission V Symposium on Image Engineering and Vision Metrology, Dresden, Germany, September 2006; pp. 266–272.
15. Rieke-Zapp, D.; Tecklenburg, W.; Peipe, J.; Hastedt, H.; Haig, C. Evaluation of the geometric stability and the accuracy potential of digital cameras-comparing mechanical stabilisation *versus* parameterisation. *ISPRS J. Photogramm.* **2009**, *64*, 248–258.
16. Yang, Y. Robust estimation for dependent observation. *Manuscr. Geod.* **1994**, *19*, 10–17.
17. Zhang, Y.; Xiong, J.; Hao, L. Photogrammetric processing of low-altitude images acquired by unpiloted aerial vehicles. *Photogramm. Rec.* **2011**, *26*, 190–211.
18. Ivory, J. An Evaluation of Photogrammetry as a Geotechnical Risk Management Tool for Open-Pit Mine Studies and for the Development of Discrete Fracture Network Models. Master's Thesis, UCL Australia, Adelaide, SA, Australia, 2012.

19. Patikova, A. Digital photogrammetry in the practice of open pit mining. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2004**, *34*, 1–4.
20. Siebert, S.; Teizer, J. Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. *Autom. Constr.* **2014**, *41*, 1–14.
21. Bemis, S.P.; Micklethwaite, S.; Turner, D.; James, M.R.; Akciz, S.; Thiele, S.T.; Bangash, H.A. Ground-based and UAV-based photogrammetry: A multi-scale, high-resolution mapping tool for structural geology and paleoseismology. *J. Struct. Geol.* **2014**, *69*, 163–178.
22. Cryderman, C.; Mah, S.B.; Shufletoski, A. Evaluation of UAV photogrammetric accuracy for mapping and earthworks computations. *Geomatica* **2014**, *68*, 309–317.
23. Holst, G.C.; Lomheim, T.S. *CMOS/CCD Sensors and Camera Systems*, 2nd ed.; JCD Publishing: Winter Park, FL, USA, 2007.
24. Luhmann, T.; Robson, S.; Kyle, S.; Harley, I. *Close Range Photogrammetry: Principles, Techniques and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
25. Brown, D.C. Close-range camera calibration. *Photogramm. Eng.* **1971**, *37*, 855–866.
26. Dörstel, C., Jacobsen, K.; Stallmann, D. DMC – Photogrammetric Accuracy – Calibration Aspects and Generation of Synthetic DMC Images. In Proceedings of Optical 3D Sensor Workshop, Zurich, Switzerland, September 2003; pp. 4–12.
27. Grewal, M.S.; Weill, L.R.; Andrews, A.P. *Global Positioning Systems, Inertial Navigation, and Integration*, 2nd ed.; John Wiley & Sons Publication: Hoboken, NJ, USA, 2007.
28. Honkavaara, E.; Hakala, T.; Markelin, L.; Rosnell, T.; Saari, H.; Makynen, J. A process for radiometric correction of UAV image blocks. *Photogramm. Fernerkund.* **2012**, *2*, 115–127.
29. Mukherjee, D.; Wu, Q.J.; Wang, G. A comparative experimental study of image feature detectors and descriptors. *Mach. Vis. Appl*. **2015**, *26*, 443–466.
30. Hirschmüller, H.; Scharstein, D. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Anal.* **2009**, *31*, 1582–1599.
31. Jung, I.L.; Chung, T.Y.; Sim, J.Y.; Kim, C.S. Consistent stereo matching under varying radiometric conditions. *IEEE Trans. Multimed.* **2013**, *15*, 56–69.
32. Adeline, K.R.M.; Chen, M.; Briottet, X.; Pang, S.K.; Paparoditis, N. Shadow detection in very high spatial resolution aerial images: A comparative study. *ISPRS J. Photogramm.* **2013**, *80*, 21–38.
33. Sona, G.; Pinto, L.; Pagliari, D.; Passoni, D.; Gini, R. Experimental analysis of different software packages for orientation and digital surface modelling from UAV images. *Earth Sci. Inform.* **2014**, *7*, 97–107.
34. Chung, K.L.; Lin, Y.R.; Huang, Y.H. Efficient shadow detection of color aerial images based on successive thresholding scheme. *IEEE Trans. Geosci. Remote Sense*. **2009**, *47*, 671–682.
35. Rumpler, M.; Daftry, S.; Tscharf, A.; Prettenthaler, R.; Hoppe, C.; Mayer, G.; Bischof, H. Automated end-to-end workflow for precise and geo-accurate reconstructions using fiducial markers. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, doi:10.5194/isprsannals-II-3-135-2014.

36. Pix4D, UAV Mapping Software. Available online: https://pix4d.com (accessed on 10 August 2015).
37. Wolf, P.R.; Dewitt, B.A. *Elements of Photogrammetry: With Applications in GIS*, 3rd ed.; The McGraw-Hill Companies: Boston, MA, USA, 2000.
38. Vanicek, P.; Krakiwsky, E.J. *Geodesy: The Concepts*, 2nd ed.; Elsevier Science Publishers: New York, NY, USA, 1986; pp. 242–283.
39. CloudCompare-Open Source project. Available online: http://www.danielgm.net/cc (accessed on 10 August 2015).
40. Bay, H.; Ess, A.; Tuytelaars, T.; van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359.

Chapter 4

4. Sparse Reconstruction- Part I

4.1 Article Presentation

4.1.1 Background

This article is part of the thesis related to objective 3.1, namely developing a technique of robust epipolar-geometry estimation and sparse matching based on evolutionary search.

Sparse matching is performed only on salient parts of stereo images, such as corner points and edges. Sparse matching techniques can be divided into two main categories of correlation-based and feature-based approaches. Feature-based approaches have gained more popularity because the recently developed feature descriptors are invariant to many sorts of geometric and photometric transformations. However, feature-based matching can be quite difficult in the presence of uniform or repetitive textures. Another important factor that affects the performance of sparse matching is the noise caused by the processes of image acquisition and feature extraction. One more serious factor is the existence of outliers, i.e. false matches among the putative correspondences (Li and Allinson, 2008; Apollonio et al., 2014). The presence of outliers is due to several factors. These factors include noisy measurements, insufficiency of local descriptors, lack of distinctive texture patterns in the scene, and the existence of repetitive features that cause high ambiguity in matching. In addition, the threshold value, which decides whether two features are similar enough to be matched, has an important impact on the amount of outliers. Therefore, outlier detection should naturally be integrated into the sparse matching procedure to avoid the outliers affecting the accuracy of epipolar-geometry estimation and structure reconstruction.

This article studies the state-of-the-art for robust sparse matching and epipolar-geometry estimation. Accordingly, a technique based on a genetic algorithm (GA) is proposed and evaluated.

4.1.2 General methodology

In order to perform robust sparse matching, first, the corresponding points need to be identified on stereo images. To this end, at this study, the scale-invariant feature transform (SIFT) detector/descriptor was used (Lowe, 2004). Afterward, the detected corresponding points (namely putative correspondences) passed through the block of genetic algorithm (GA). The outputs of this block were the elements of epipolar geometry (defined by a fundamental matrix) and a set of initial inliers (called the inlier set of minimum cardinality). These elements, then, passed through a block of inlier-classification, which was aimed at detecting all the inliers among putative correspondences. This was performed by computing the uncertainty of the estimated fundamental matrix and its effect on the residuals of inlier matches. Figure 4.1.1 presents the general methodology of this article.

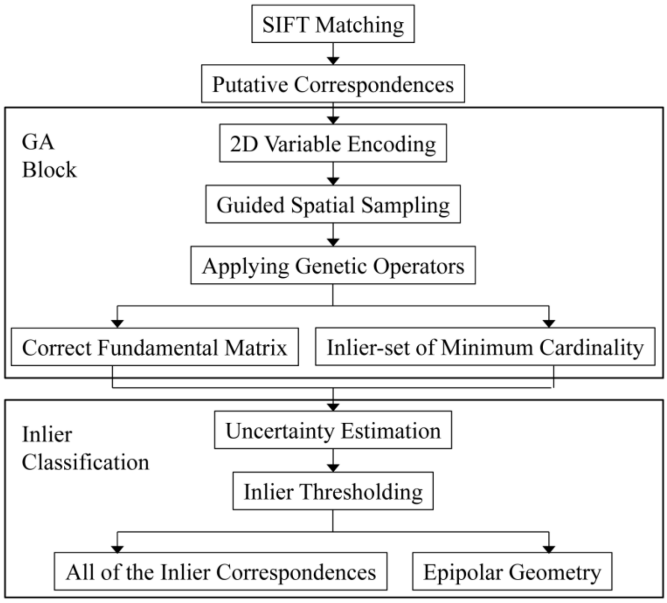


Figure 4.1.1. Methodological flowchart of robust sparse matching and epipolar-geometry estimation

Evolutionary Optimization for Robust Sparse Matching and Epipolar-Geometry Estimation

by Mozhdeh Shahbazi, Gunho Sohn and Jérôme Théau

Computer Vision and Image Understanding,

ISSN: 1077-3142

New Submission

Abstract

In this paper, a robust technique based on a genetic algorithm is proposed for estimating two-view epipolar-geometry of un-calibrated perspective stereo images and detecting inliers from putative correspondences containing a high percentage of outliers. The proposed technique aims at increasing computational efficiency and accuracy in comparison with the state-of-the-art via the following contributions: i) replacing random search with evolutionary search applying new strategies of encoding and guided sampling; ii) robustly estimating the epipolar geometry via detecting a more-than-enough set of inliers without making any assumptions about the probability distribution of the residuals; iii) detecting all the inliers by considering the uncertainty of the estimated model. The proposed method was evaluated both on synthetic data and real images. The results were compared with the most popular techniques from the state-of-the-art, including RANSAC, MSAC, MLESAC, Cov-RANSAC, LO-RANSAC, PROSAC, RRANSAC-SPRT, StaRSAC, Multi-GS RANSAC and LMedS. Experimental results showed that the proposed approach performed better than other methods in terms of the accuracy of inlier detection and epipolar-geometry estimation as well as the computational efficiency for datasets majorly contaminated by outliers and noise.

Resumé

Dans cet article, une technique robuste basée sur un algorithme génétique est proposée pour estimer la géométrie épipolaire des images stéréo et pour détecter les « inliers » parmi des correspondances putatives contenant un pourcentage élevé d'erreurs aberrantes. La technique proposée vise aussi à accroitre l'efficacité de calcul et la précision par rapport à l'état de l'art via les contributions suivantes: i) le remplacement de la recherche aléatoire par la recherche évolutive appliquant de nouvelles stratégies de codage et d'échantillonnage, ii) l'estimation du modèle de géométrie robuste via la détection d'un ensemble de « inliers » sans hypothèse préalable sur la distribution de probabilité des résidus, iii) la détection de tous les « inliers » en considérant l'incertitude du modèle estimé. La méthode présentée a été évaluée à la fois sur des données synthétiques et des images réelles. Les résultats ont été comparés avec les techniques les plus populaires de l'état de l'art incluant : RANSAC, CCSM, MLESAC, Cov-RANSAC, LO-RANSAC, PROSAC, RRANSAC-SPRT, StaRSAC, Multi-GS RANSAC et LMedS. Les résultats expérimentaux ont montré que l'approche proposée est plus performante que d'autres méthodes en termes de précision de détection des « inlier », d'estimation de la géométrie épipolaire et d'efficacité de calcul pour les données contaminées par des valeurs aberrantes et du bruit.

Key Words: Sparse Matching, Outlier Detection, Genetic Algorithm, Epipolar Geometry, Evolutionary Search, Guided Sampling, Adaptive Thresholding

4.2 Introduction

Sparse image matching is the main stage in many computer vision applications, including structure from motion (SfM) and robotic navigation. In contrast to dense image matching, where image correspondences are established at nearly each pixel, sparse matching establishes the correspondences at salient image points only. Recent researches apply sparse matching to address a variety of problems including simultaneous localization and mapping (Schmidt, 2014; Hartmann et al., 2013), feature tracking (Wendel et al., 2012), dense image matching (Wöhler, 2013) and real-time mosaicing (Kekec et al., 2014). The results of sparse matching are usually contaminated with a considerable percentage of false correspondences. This is due to several factors, which include noisy measurements, the inefficiency of local descriptors, lack of texture diversity, and the existence of repetitive features that cause matching ambiguity (Li and Allinson, 2008; Apollonio et al., 2014). Therefore, outlier detection should essentially be integrated into sparse matching. From now on, the term *putative correspondence* is used for referring to the raw results of sparse matching. Accordingly, the term *inlier* refers to true matches among the putative correspondences and the term *outlier* refers to false matches. Besides, the terms *correspondence* and *match* have equal meanings throughout this paper.

Generally, outlier detection techniques are based on the fact that inliers have some spatial characteristics in common. The correspondences that are not consistent with such spatial characteristics can be classified as outliers. The idea of using epipolar geometry as a constraint to detect inliers/outliers has been proposed in several studies. With this regard, the matching problem turns into two problems of i) robust estimation of epipolar geometry from putative correspondences, ii) detecting the entire set of inliers using the estimated model[1]. In Section 4.3, a review of such techniques is presented. Figure 4.2.1 presents the summary of the inlier detection techniques that are discussed in Section 4.3.

In this paper, we also focus on the problem of detecting inliers/outliers from putative correspondences based on the robust estimation of epipolar geometry. To this end, we adapt the integer-coded genetic algorithm (GA). The proposed technique can be considered as an extension and generalization of RANSAC-like methods that has the following distinctive characteristics. First, random sampling is replaced with the evolutionary search. The evolutionary search speeds up the algorithm by bringing two major advantages: the search begins from a population of sample sets instead of a single initialization, and new sample sets are generated considering the feedback information obtained by evaluating previous sample sets. Second, a guided sampling scheme based on the spatial distribution of the correspondences is proposed and applied to the evolutionary search. This sampling scheme increases the robustness of the solutions against degenerate configurations and local optima without requiring additional computation or prior information about the matches. Third, the objective function of robust estimation is not defined based on the support cardinality, and the robust least teamed sum of squared residuals is used instead. Therefore, there is no need to successively define a threshold at any iteration for detecting the support of the estimated model, or to assume any specific probability distribution for outlier/inlier residuals. Finally, to identify all the inliers, a classification method based on adaptive

[1] From now on, the term *model* refers to the two-view epipolar geometry, which is described here by the fundamental matrix.

thresholding is proposed, as opposed to using a fixed threshold. In this method, the uncertainty of the estimated model is taken into account to identify all the inliers correctly.

The rest of the paper is organized as follows. The main problem of simultaneous inlier-detection and epipolar-geometry estimation is formulated in section 4.4. Section 4.5 describes the solution using the genetic algorithm, which is followed by the method of detecting all the inliers. The experimental results are discussed in section 4.6, and the conclusion is presented in section 4.7.

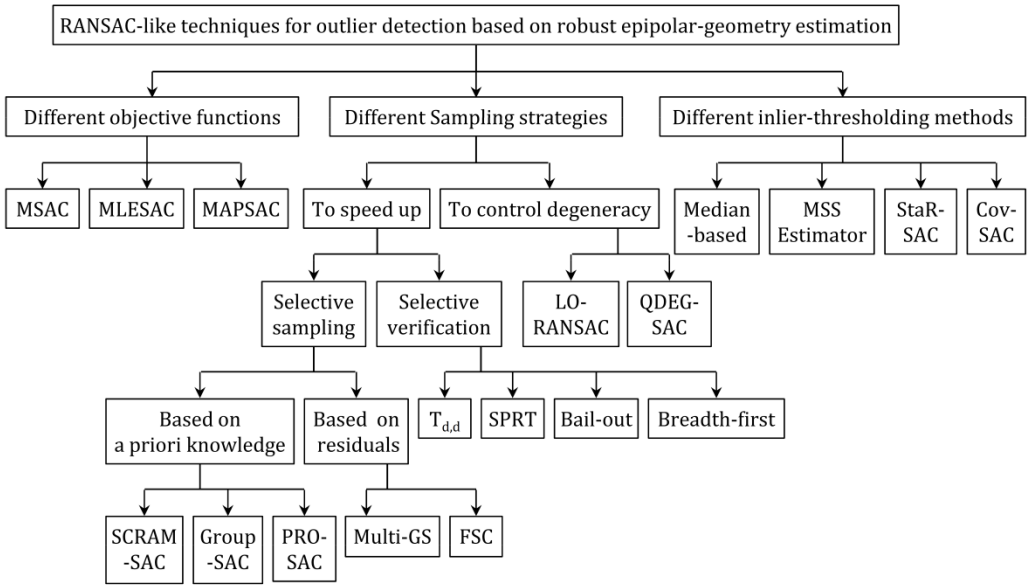


Figure 4.2.1. Summary of outlier detection techniques in stereo sparse matching based on robust estimation of epipolar geometry

4.3 Related Work

Random Sample Consensus (RANSAC) techniques are most popular approaches in fields of robust estimation. Basically, RANSAC aims at determining the optimal model from an outlier-free sample set of correspondences by maximizing the support size of the model. The inliers that support the model are detected as matches whose residuals from the estimated model are less than a given threshold. To find an outlier-free sample set, successive random sampling is performed. In order to ensure, with probability p, that at least one outlier-free set of m matches is sampled from a dataset containing ε percent of inliers, at least k sample sets should be drawn such that $k \geq \log(1-p)/\log(1-\varepsilon^m)$. This also means that within RANSAC, approximately, $\log(1-p)^{-1}$ good models are generated before termination.

By this definition, six major questions are involved in RANSAC-like techniques. i) Is maximizing the support cardinality a robust objective function when no information about the rate of the outliers is available? ii) How to handle a large number of required samples in cases where ε is very small? iii) Why the algorithm does not take advantage of the probability of generating good hypotheses before reaching the termination criterion; i.e. why not including the feedback of previous samples in the

sampling procedure? iv) How the robustness of the estimated model can be ensured against the influence of noise since it is relying on a minimal (just enough) subset of inliers? v) How to control the effect of degenerate sample sets which naturally maximize the support cardinality? vi) Does the threshold used to detect the inliers reflect the uncertainty of the estimated model as well? Some of these questions are answered by different variants of RANSAC, which are discussed here.

Unlike the standard RANSAC, there are improved variants, which use robust objective functions to determine the support cardinality. In m-estimator sample consensus (MSAC), inliers are scored based on their fitness and outliers are scored with a non-zero constant penalty (Torr and Zisserman, 1998). Maximum likelihood estimation sample consensus (MLESAC) maximizes the log-likelihood of the solution via RANSAC process by assuming that outliers are distributed uniformly and residuals distribute a Gaussian function over inliers (Torr and Zisserman, 2000). MAPSAC is also a refined version of MLESAC with Bayesian parameter estimation (Torr, 2002). These objective functions make models with similar inlier scores more distinguishable. However, they make certain assumptions about the distribution of the residuals either for inliers or outliers. Besides, they still score the models and detect the final inliers by applying a hard threshold to the residuals. Generally, this threshold is determined from the standard deviation of the residuals themselves. Assuming that the noise in data-points follows a Gaussian distribution $\mathrm{N}(0,\sigma)$, then the residuals ideally follow a Chi-square distribution. Therefore, the threshold can be expressed as $\chi^{-1}(p)\sigma^2$, where $\chi(p)$ is the cumulative Chi-square distribution with one degree-of-freedom at probability p as the fraction of the inliers to be captured (e.g. 0.95). However, this assumption is true by ignoring the uncertainty of the model itself. Besides, estimating the standard deviation σ at any RANSAC iteration is another problem. One of the most common methods is to estimate this variable from the median of the residuals on the potential inliers that support the best tentative model. Therefore, it might not be robust against outliers (Torr and Murray, 1997). Another method is to determine a Gaussian distribution fitting to the smallest residuals in the dataset (modified selective statistical estimator (MSSE) by Bab-Hadiashar and Suter, 1999). While this method is more robust against outliers, it is sensitive to the distinction of inliers from outliers. One strategy to eliminate the requirement of a fixed threshold is to run RANSAC several times using a range of pre-determined thresholds (StaRSAC method by Choi and Medioni (2009)). However, depending on the range of the thresholds to be tested and the number of RANSAC executions, this strategy can be computationally exhaustive. In Cov-RANSAC algorithm, the uncertainty of the model estimated from the minimal sample set is used to determine a subset of potential inliers, to which the ordinary RANSAC is applied afterward (Raguram et al., 2009). However, the uncertainty of the estimated model highly depends on the spatial configuration and noise-level of the points appearing in the minimal sample set.

The sampling strategy in RANSAC is also an important factor as it influences the algorithm efficiency with respect to both the number of RANSAC iterations and degeneracy of the estimated model. There are several methods to control each of these two factors. The main contributions with this regard are discussed here. To control the speed of the algorithm, two strategies can be applied. The first strategy is to enforce an initial consistency check to filter the putative matches. This consistency can be measured

as the fraction of neighboring features in a region around a point in one image whose correspondences fall into the similar region in the other image (SCRAMSAC method by Sattler et al. (2009)). This strategy is sensitive to the region size and the threshold used to define the spatial consistency. Alternatively, samples can be drawn by an order based on the matching scores (PROSAC method by Chum and Matas (2005)). However, such a strategy is not effective when foreground motion happens. In addition, the scenes with repetitive textures may result in many false matches with high matching scores. Another example of such strategies is to assume that matches have a natural grouping structure, in which some of the groups have higher inlier probability than others (GroupSAC method by Ni et al. (2009)). However, finding a meaningful and efficient grouping among the matches is itself a concerning challenge in different applications. In resume, the strategies mentioned so far mostly require supplementary information about the scene or the matches. Comparatively, guided sampling based on the information from sorted residuals can be used to accelerate the hypothesis generation while avoiding any application-specific ordering or grouping technique (Multi-GS by Chin et al., 2012). In this method, sampling is guided towards selecting the points that are rising from the same structure. This strategy speeds up the procedure to reach an outlier-free sample set. However, this method causes more computational complexity, since every point in the hypothesized sample set should be compared against all the other points in the dataset in order to determine its intersection (in terms of structure) with them. The fast consensus sampling (FCS) method based on the residuals is developed by Cheng and Lai (2009). In this method, proposal probabilities are calculated for matches based on their normalized residuals and a concentration score. Although this method accelerates the sampling, it is still sensitive to degeneracy, image noise, and uncertainty of the model estimation. This is due to the fact that it reduces the number of potential inliers by thresholding the proposal probabilities that are, themselves, dependent on the robust estimation of normalized residuals. GASAC is also another technique which is, to the best of authors' knowledge, the most similar technique to the one proposed in our study. The difference of GASAC with typical RANSAC is that the random sampling is replaced with evolutionary search based on classic genetic algorithms (Rodehorst and Hellwich, 2006). This study shows that the computational cost could be speed up 13 times by applying the evolutionary search instead of the random one. However, the technique is applied to small datasets with less than 50% outliers, the objective function is still based on support cardinality, and it introduces no solution to avoid local optima (such as degeneracy). The second strategy proposed in the literature in order to speed up RANSAC is to reduce the solution space by only verifying the hypotheses with a higher probability of being optimal. These high-probable hypotheses can be selected by $T_{d,d}$ test (Matas and Chum, 2004), bail-out test (Capel, 2005) or sequential probability ratio test (SPRT) (Matas and Chum, 2005; Chum and Matas, 2008). In addition, the hypothesis verification can be performed preemptively in a breadth-first manner only for a fixed number of sample sets (Nistér, 2005). These techniques may increase the number of required hypotheses as good models may wrongly be rejected by not being verified completely.

Degeneracy in robust epipolar-geometry estimation occurs when one/more degenerate configurations exist in the scene. This usually happens when the majority of matches belong to a dominant plane in the scene and the rest of the matches are not on the plane (planar degeneracy), or when the matches belong to a very small region of an image (ill configuration). It has been observed that, in these cases,

RANSAC-like algorithms result in a model with a very large support while it is completely incorrect. This behavior can be explained by the fact that a high inlier support can be obtained even if the sample set includes some outliers and at least five inliers that belong to a dominant plane or to a very small area of the image. This large support causes the termination of RANSAC before a non-degenerate outlier-free sample set can be picked up (Chum et al., 2005). The main strategy to control the degeneracy of solutions is to re-investigate the support of the best tentative model either locally or globally. For instance, the support of the best model can be re-sampled and the model estimated from those subsamples can be compared to the best one to find out the degenerate models (LO-RANSAC method by Lebeda et al. (2012)). Another example of such strategies is the QDEGSAC (proposed by Frahm and Pollefeys (2006)), where a hierarchical RANSAC is performed by changing the number of the parameters in the model and verifying it over the entire dataset. The main issue concerned with these techniques is that they cause additional operations, as a separate mechanism is added to the original RANSAC; i.e. they do not directly handle degeneracy in the sampling process. Universal RANSAC (USAC) is also a modular fusion of some of the mentioned RANSAC algorithms including PROSAC sampling, SPRT verification and LO-RANSAC local optimization (Raguram et al., 2013). In general, its performance is better than any of the single modules integrated into the universal implementation. However, it does not improve any of the modules individually.

4.4 Problem Formulation

Notations: Column vectors are represented by italic, bold lower case letters, such as $\boldsymbol{x}$. Therefore, $\boldsymbol{x}^{\mathrm{T}}$ (the transpose of $\boldsymbol{x}$) is a row vector. Matrices are denoted by italic uppercase letters, such as F. The elements of the matrix are denoted as F_{ij}, where i represents the row, and j represents the column number. Sets are denoted by italic, bold uppercase letters, such as $\boldsymbol{U}$.

4.4.1 Fundamental theories: Two-view epipolar geometry

Epipolar geometry defines the geometry of stereo vision, all elements of which can be captured by a matrix called the fundamental matrix. It can also be defined by an essential matrix in case of calibrated images, where the parameters of intrinsic camera calibration are known. There are different methods for estimating the fundamental matrix: linear, iteration-based, and robust techniques (Armangué and J. Salvi, 2003). Robust techniques, which use linear techniques as their base, are the most applicable ones since they can handle the presence of outliers and, finally, can detect the inlier matches required for structure reconstruction. This category of estimation is considered in this study. The following paragraphs present the theoretical background with this regard.

For any pair of homogenous coordinates of correspondences $\boldsymbol{x} \leftrightarrow \boldsymbol{x}'$, the fundamental matrix (F) can be described as follows.

$$\boldsymbol{x}'^{\mathrm{T}} F \boldsymbol{x} = 0 \tag{1}$$

From Equation (1), it can be noticed that F is defined up to an unknown scale. It is also a rank-2 matrix with zero determinant. Consequently, it has only seven degrees of freedom (Hartley and Zisserman, 2003).

Given m matches, it would be possible to form a homogeneous system of m linear equations from Equation (1) as:

$$\boldsymbol{U}(\boldsymbol{f},\boldsymbol{x},\boldsymbol{x}') = A\boldsymbol{f} = 0, \tag{2}$$

where $\boldsymbol{f}$ is the vector containing the nine elements of matrix F, and A is the $m\times 9$ coefficient matrix. The coordinates of the points are usually normalized so that A can gain a better condition number (Hartley, 1997).

In robust estimation techniques, non-parameterized linear methods of fundamental-matrix estimation are used due to the reasonable agreement between the speed and accuracy yielded from these methods (Zhong et al., 2006). Given exactly seven matches (m=7), it would be possible to determine $\boldsymbol{f}$ by spanning the two-dimensional nullspace of A and applying the rank deficiency constraint to it (seven-point algorithm). However, this results in up to three fundamental matrices. Thus, when the goal is the robust estimation of the fundamental matrix, computational expenses for hypothesis evaluation would increase up to three times. Depending on the total number of correspondences and percentage of outliers, this can be a major drawback. Therefore, eight or more matches ($m\geq 8$) are required to determine a single solution for the fundamental matrix (8-point algorithms). Using at least eight points, the solution to F can be found from Equation (2) by linear least-squares methods. At the end, the rank deficiency constraint must be applied to the estimated F by setting its smallest singular value to zero (Hartley, 1997).

In the case of the 8-point algorithm, an additional constraint should be imposed to define an arbitrary scale factor for F and to prevent the trivial solution $F=0$. There are two options with this regard. One would be to fix the 2-norm of the fundamental matrix (e.g. $\|\boldsymbol{f}\|=1$). The other would be to fix one element of the matrix (e.g. $F_{33}=1$). In the first case ($\|\boldsymbol{f}\|=1$), an orthogonal least-square minimization should be applied. It can be shown that the solution is the right singular vector of A corresponding to the smallest eigenvalue of $A^{\mathrm{T}}A$, which can be determined by singular value decomposition (SVD) of matrix A. Technically, to compute the SVD of the matrix $A\in\mathbb{R}^{m\times 9}$, 8019+162$m$ flops (amount of arithmetic operations) are at least required (Golub and Van Loan, 2012). In the second case ($F_{33}=1$), a set of non-homogenous linear equations, with equation matrix $A\in\mathbb{R}^{m\times 8}$, will be resulted, solving which requires 170+64m flops (Golub and Van Loan, 2012). Therefore, applying the linear scale constraint ($F_{33}=1$) is computationally 13 times faster than the non-linear constraint for the minimal case of m=8. In addition, it would be possible to add the observation weights directly and use weighted linear least-squares techniques (Zhang, 1998). However, using this option is quite risky as it influences the estimation of F if the coefficient F_{33} approaches zero (Zheng et al., 2013). Such situations ($F_{33}\to 0$) can be raised in cases that we will call *poor* camera motion models. In these cases, the above-mentioned assumption ($F_{33}=1$) fails; e.g. the rotation of the second camera coordinate system with respect to the first one is mainly

planar, the camera motion contains pure translations or two cameras are shifted mainly along each other's optical axes and other translation elements are zero. To avoid such exceptions, one way would be to examine all the nine elements of the fundamental matrix for setting them to a constant non-zero value (i.e. F_{ij}=1) and choose the best solution (Chesi et al., 2002). However, this increases the computational time.

In this study, Gaussian elimination with partial pivoting is used to detect the free variable of the consistent linear system of Equation (2) and to solve it. The free variable corresponds to the element of the fundamental matrix, whose corresponding column in the coefficient matrix (A) is not a pivot column. Therefore, that element cannot be zero and can take a fixed value, e.g. 1, to resolve scale deficiency of the fundamental matrix. In terms of complexity, partial pivoting requires $2/3m^3$ flops. Therefore, its application is not efficient for large datasets. However, in the case of robust estimation where only a minimal number of 8 points is used, this method requires only 341 flops (compared to 9351 flops for SVD decomposition).

Given an estimated fundamental matrix, it is possible to express how well the matches fit to it by calculating the residuals of the matches (Torr and Murray, 1997). There are various error measures to represent the residuals including algebraic distance, epipolar-weighted distance, 2-sided point to epipolar line distance and Sampson distance (Equation (3)).

$$d_S = \left|\boldsymbol{x}'^T F \boldsymbol{x}\right| / \sqrt{l_1^2 + l_2^2 + l_1'^2 + l_2'^2} \tag{3}$$

where:

$[l_1, l_2, l_3]^T = F\boldsymbol{x}$ and $[l_1', l_2', l_3'] = \boldsymbol{x}'^T F$

The ideal measure for robust estimation should be robust to image noise. That is, the residuals should not increase unexpectedly by increasing the noise magnitude to some small degrees. As the previous studies have shown, Sampson distance is less sensitive to image noise in comparison with other error measures. Therefore, Sampson distance is used in this study to represent the residuals.

4.4.2 Robust estimation problem

The main problem of robust estimation is to find a minimal sample set of inliers, from which the fundamental matrix (the model) can be estimated *correctly*. To determine how *correct* an estimated model is, an objective function is required. In RANSAC robust estimation, this objective function is defined as the support of the estimated model. However, as mentioned in section 4.3, this requires a threshold to decide whether a correspondence supports the estimated model or not. In this study, in order to avoid such a threshold, the concept of least trimmed squares (Hubert et al., 2008) is applied. The objective is to minimize the sum of squared residuals over a minimum number of inliers, which we would call *inlier set of minimum cardinality*. This minimum cardinality could be easily hypothesized without loss of generality. For instance, considering a data set with 60% outliers, one can assume that at least 10% of the data are inliers without any knowledge of the errors. Generally, the minimum number of inliers would be a more relaxed assumption in comparison with the approximate ratio of outliers.

To formulate the problem, we assume that at least n^* matches in the data set of size n are certainly inliers. In fact, we set the breakdown point of the inlier detection to $(n-n^*)/n$. Then, the cost of the fundamental matrix is measured as the sum of squared Sampson distances over an inlier set of size n^*. Therefore, the optimal fundamental matrix, $\hat{F}$, can be found as in Equation (4), where d_i is the Sampson distance computed using Equation (3) for i^{th} correspondence ($\boldsymbol{x_i} \leftrightarrow \boldsymbol{x_i'}$).

$$\hat{F} = \operatorname{argmin}\left(\sum_{i=1}^{n^*} d_i^2\right) \qquad (4)$$
$$\text{s.t. } d_1^2 < d_2^2 < \cdots < d_{n^*}^2 < d_{n^*+1}^2 < \cdots < d_n^2$$

4.4.3 Robust Estimation via Genetic Algorithm

The robust estimation problem (section 4.4.2) can be considered as an integer, nonlinear problem where the variables are sets of integer labels representing the putative correspondences, and the objective function is nonlinear (Equation (4)). There are several ways to solve such a problem, including dynamic programming, mixed-integer programming, and heuristic optimization. Although dynamic programming and mixed-integer, nonlinear programming provide good tools to address such problems, they still require considerable computational effort (Elbeltagi et al., 2005). In contrast, heuristic and meta-heuristic optimization techniques are capable of finding good-enough solutions within a reasonable time by applying probabilistic transition rules rather than deterministic ones. Accordingly, in this paper, a modified version of the integer-coded genetic algorithm, originally proposed by Deep et al. (2009), is applied. Basically, GA encodes the decision variables (chromosomes) into a cellular environment. Then, several sub-sets of the chromosomes are sampled; each sample set is called an individual. Several individuals form the population. In any iteration of the evolution, the parent individuals in the population are evaluated, and the genetic operators are performed on them to produce a new population. This evolution continues the same way until reaching an optimal solution.

Terminology: Each corresponding point is a chromosome. A minimal sample set of correspondences is an individual.

The overall pseudo-code of the proposed technique is shown in Table 4.5.1.

Table 4.5.1. The pseudo-code of the proposed robust estimation technique via genetic algorithm

Input
Putative correspondences
Output
Estimated epipolar geometry (fundamental matrix) and the entire set of inlier matches
a) Genetic Algorithm

- Input: The label set of matches

- Output: The accurate fundamental matrix (F) and an inlier set of minimum cardinality

1. Initialize the first population by guided sampling

While the best solution is improved

2. Compute F from each individual

3. Evaluate each individual by computing the sum of the n^* smallest squared Sampson residuals

4. Perform genetic operators on the individuals of the current population and reproduce the next population

5. Save the best solution so far with the inlier set of minimum cardinality (size= n^*) associated with it

End while

6. Re-estimate the fundamental matrix ($\hat{F}$) using the inlier set of minimum cardinality from the best solution ($\hat{\boldsymbol{I}}$)

b) Estimate the uncertainty of the model ($\hat{\Sigma}_{\hat{F}}$)

c) Estimate the average and uncertainty of the Sampson residuals (μ_d^*, σ_d^*) for matches belonging to $\hat{\boldsymbol{I}}$ to determine the outlying threshold

d) Compute and threshold the Sampson residuals on other matches to identify the entire set of inliers

4.4.4 Encoding GA variables

Since GA operators take the integers as input to create new solutions, these integers should represent the matches geometrically. To this end, in this study, every match is labeled based on its position on the reference (first) image: one integer for its horizontal and one integer for its vertical position relative to other points. Therefore, the whole set of matches could be represented by a label set as $\mathrm{S} = \{k(i,j) \,|\, 1 \le k \le n,\ 1 \le i \le w,\ 1 \le j \le h\}$, where w and h represent the vertical and horizontal extent of the matches, and n is the total number of matches. Figure 4.5.1 shows a naive example of a label set for a few matches. The empty cells in the label set are filled with nearest non-empty cell values. The following examples justify our choice of encoding scheme.

From the example given by Figure 4.5.1, assume that we have a chromosome (i=6, j=7), where i shows the horizontal position of the match on the left image, and j shows its vertical position. Using the label set (***S***), we can determine to which match this position is referring. In this case, it is referring to match number 5. This encoding of the matches allows us to define the genetic operators on the coordinates of the matches rather than their permuted labels, which is of course geometrically more meaningful. As an example, consider the mutation operator whose job is to locally explore the solutions. For instance,

assume that the mutation operator is applied to chromosome (6, 7), and it is mutated to (7, 8). From the label set, we can notice that (7, 8) refers to the match number 1; i.e. the 5th match should be replaced with the 1st match. This mutation operator totally makes sense, as one match is replaced with another match explored in its local neighborhood. However, if a random permutation of integers, from 1 to 7, were used for encoding the matches, then the mutation of number 5 could result in number 6. That is the 5th match at position (6, 7) should have been replaced with the 6th match at position (141, 42). It can be noticed that the results of such mutation does not geometrically make sense, as it mutates the matches randomly but not locally at all.

Part of the label-set (***S***)

	6	7	8	9	10	11
6	5	5	3	3	4	**4**
7	**5**	5	3	3	4	4
8	5	1	3	**3**	3	3
9	1	**1**	1	3	3	3
10	1	2	**2**	2	2	3

List of some of the putative matches in dataset

Match number (*k*)	Horizontal integer position (*i*)	Vertical integer position (*j*)
1	7	9
2	8	10
3	9	8
4	11	6
5	6	7
6	141	42
7	57	23

Figure 4.5.1. Example of a label set for encoding the matches

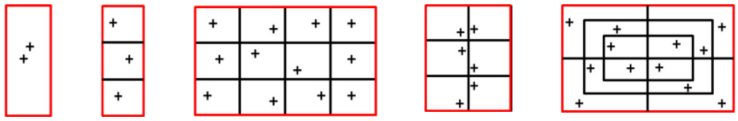
4.4.5 Sampling GA individuals

The first step to initiate GA is to generate a random population of individuals. As mentioned in section 4.4.1, at least 8 points are required to form a minimal sample set from which the fundamental matrix can be calculated. As discussed in section 4.3, random sampling can be either done uniformly or with an order based on the quality of matches. However, none of these sampling strategies is robust to degenerate configurations. To avoid the degenerate solutions (also local optima), guided sampling based on the spatial distribution of matches is proposed in this paper. Using this strategy, first, distribution of matches all over the overlapping area of the stereo images is guaranteed. Secondly, the density of matches at different regions of the images is taken into account. To clarify our reason for considering the spatial distribution of the matches, consider the following example.

Assume that we are supposed to randomly pick up two points at the rectangle of Figure 4.5.2a, aiming to provide a reasonable horizontal and vertical distance between the points. Evidently, there is a high risk of selecting two points too close to each other. A solution in order to decrease such a risk is to divide the rectangle into three equal-area regions and to pick up three points instead of two (Figure 4.5.2b). Now imagine that the rectangle of Figure 4.5.2a represents one-fourth of the overlapping area between two stereo images. Extending the same reasoning, we can pick up a sample set of 12 points from a 4×3 grid instead of 8 uniformly sampled points (Figure 4.5.2c) to ensure a better spatial configuration. Evidently, more than 12 grids could be considered for sampling. However, the more the number of the points, the higher the risk of encountering outliers. Besides, 12 data points were already proposed by other studies such as Chum et al. (2003) and Lebeda et al. (2012).

Of course, there is still a probability that the sampled points be too close to each other. Such cases might especially happen when the points are selected from the regions close to same edges and corners of the grid cells. For instance, consider a half of the overlapping area, from which six points are to be picked up (Figure 4.5.2d).[1] We aim at calculating the probability of sampling 6 ill-configured points from the areas close to corners of the grid cells. Two criteria are considered to decide if a set of six points is ill-configured: i) the surface area of the convex hull formed by these points is less than 1/3 of total area, ii) the distance between the two furthest points in the set is less than half of the dimension of the whole rectangle at both horizontal and vertical directions. In total, there are 4^6 possibilities for picking up a sample set of 6 points from the corner areas. Among these possibilities, 39% of the sample sets are ill-configured with respect to the area criterion and 28% of the sample sets are ill-configured according to the distance criterion.

In this study, a new gridding method is applied to divide the overlapping area into 12 regions of equal area, which reduces the chance of picking up ill-configured samples (Figure 4.5.2e). This gridding scheme can reduce the chance of picking up points too close to each other in any direction. The same analysis was performed for this grid. There is a total of $6^4 \times 4^2$ possibilities for picking up six points near the corners. Among these cases, 36% of the sample sets are ill-configured regarding the area criterion and 14% according to the distance criterion. Therefore, the risk of sampling ill-configured points using this method is up to two times less than the regular grid-based strategy.

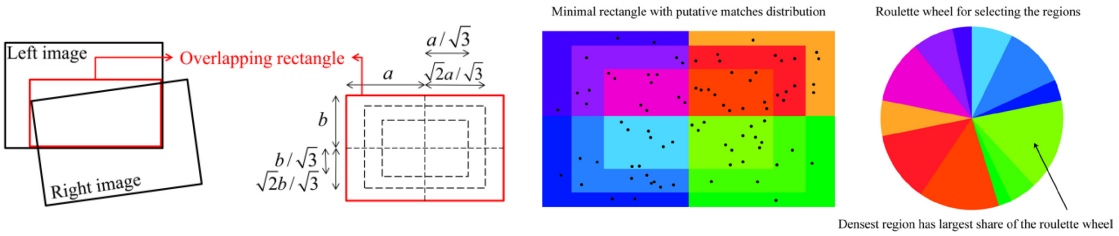


a b c d e

Figure 4.5.2. Different examples of sampling strategies to avoid degenerate and ill-configured configurations

Therefore, to guide the sampling based on spatial distribution, the overlapping rectangle containing the model area on the first image is divided into 12 sub-regions of equal area as in Figure 4.5.3a. This overlapping area is simply estimated via the extent of the putative matches on the first image. Then, the density of each region is calculated as the number of matches enclosed by it, normalized by the total number of matches. For the first half of the GA population, every individual (minimal sample set) is made of twelve matches that are randomly picked up from twelve different regions. For the other half of the population, the matches are picked up from twelve regions that are selected successively in roulette-wheel selection. The density of a region determines its probability to participate in sampling. Figure 4.5.3b illustrates an example of putative matches distributed on the overlapping rectangle and the roulette wheel corresponding to it.

[1] Although the whole overlap area could be analyzed, the half of the zone is selected for simplicity. The general conclusions are, however, applicable to the whole area.



a b

Figure 4.5.3. Guided sampling: (a) finding the overlapping rectangle containing the model area (from the extent of putative matches) and dividing it to sub-regions of equal area, (b) sampling based on the density of the regions

Accordingly, a minimal sample set for estimating the fundamental matrix can be represented as a set $\boldsymbol{\Omega} = \{m_1,..,m_{12} \mid \underset{\forall i \in \{1:12\}}{m_i} \in \boldsymbol{S} \text{ and } \underset{\forall i \in \{1:12\}}{m_i} \neq \underset{\forall j \in \{1:12\}, j \neq i}{m_j}\}$. Every match, m_i, is composed of corresponding points $\boldsymbol{x}_{m_i} \leftrightarrow \boldsymbol{x}'_{m_i}$. Once the population is formed, for every set $\boldsymbol{\Omega}$ in the population, the matches should be substituted into Equation (5) to determine the fundamental matrix.

$$\boldsymbol{U}(F,\boldsymbol{\Omega}) = \boldsymbol{x}'^{T}_{m_i} F\, \boldsymbol{x}_{m_i} = 0 \quad , \quad i \in \{1:12\} \tag{5}$$

Then, each fundamental matrix is evaluated using the cost function in Equation (4), and a fitness value is assigned to it; the lower the value of the cost function, the *fitter* the solution.

4.4.6 Genetic operators

Once the individuals are evaluated, a selection operator is applied to the population to allocate the instances of fitter individuals for entering a mating pool as parents to reproduce a new generation. The tournament selection is used here due to its higher computational efficiency over other selection techniques (Deep et al., 2009). Afterward, new individuals are generated from the selected parents by crossover and mutation operators during a reproduction process.

The crossover operator combines the chromosomes to explore the search space. It re-forms the previously sampled points to a new configuration. Therefore, if two sample sets contain ill-configured matches, the crossover operation increases the chance of reaching a better spatial configuration from the new combination of such points. Mathematically, two offsprings, $\boldsymbol{\Omega}_{offspring1}$ and $\boldsymbol{\Omega}_{offspring2}$, are created by combining two selected parents, $\boldsymbol{\Omega}_{parent1} = \{m_i^{par1} \mid \mathrm{i} \in \{1:12\}\}$ and $\boldsymbol{\Omega}_{parent2} = \{m_i^{par2} \mid \mathrm{i} \in \{1:12\}\}$, using the crossover operator as in Equation (6).

$$\begin{aligned} &\boldsymbol{\Omega}_{offspring1} = \{m_i^{offs1} \mid \mathrm{i} \in \{1:12\}\} \quad , \quad \boldsymbol{\Omega}_{offspring2} = \{m_i^{offs2} \mid \mathrm{i} \in \{1:12\}\} \\ &\text{where:} \\ &m_i^{offs1} = round(m_i^{par1} + \beta_i \left| m_i^{par1} - m_i^{par2} \right|) \, , \quad m_i^{offs2} = round(m_i^{par2} + \beta_i \left| m_i^{par1} - m_i^{par2} \right|) \end{aligned} \tag{6}$$

Instead of using the original Laplace-distributed parameter proposed by Deep et al. (2009) for β_i, we define β_i slightly differently via Equation (7) in order to respect the geometrical extension of matches on the images. In Equation (7), $u_i \in [0,1]$ is a uniformly distributed random number.

$$\beta_i = \max\left(\frac{1-m_i^{par1}}{\left|m_i^{par1}-m_i^2\right|}, \frac{1-m_i^2}{\left|m_i^{par1}-m_i^2\right|}\right) + \left(\min\left(\frac{(w,h)-m_i^{par1}}{\left|m_i^{par1}-m_i^{par2}\right|}, \frac{(w,h)-m_i^{par2}}{\left|m_i^{par1}-m_i^{par2}\right|}\right) - \max\left(\frac{1-m_i^{par1}}{\left|m_i^{par1}-m_i^{par2}\right|}, \frac{1-m_i^{par2}}{\left|m_i^{par1}-m_i^{par2}\right|}\right)\right) u_i \quad (7)$$

While crossover is conducted on two parental individuals, mutation is carried out on one individual. The mutation operation takes into account the fact that usually inliers tend to be closer to each other than outliers (as assumed in NAPSAC robust estimation method by Myatt et al. (2002)). Therefore, it performs a local search for finding more inliers around currently sampled points. A mutated solution, $\boldsymbol{\Omega}^{mut}$, is created from a parent solution $\boldsymbol{\Omega}^{par}$ by the following equation (Deep et al., 2009).

$$\Omega^{mut} = \begin{cases} \left[\Omega^{par} - s(\Omega^{par} - \min(\Omega^{par}))\right], \tau < u \\ \left[\Omega^{par} + s(\max(\Omega^{par}) - \Omega^{par})\right], \tau \geq u \end{cases} \quad (8)$$

In Equation (8), $\tau = \frac{\boldsymbol{\Omega} - \min(\boldsymbol{\Omega})}{\max(\boldsymbol{\Omega}) - \min(\boldsymbol{\Omega})}$, u is a random uniformly distributed number between zero and one, and $s = (s_1)^2$ follows a power distribution with s_1 having a random value between zero and one.

At any iteration, random exploration is also performed by generating a fixed number of individuals based on the guided sampling strategy. Generating random solutions as a fixed portion of the population reduces the chance of converging to local optima.

The replacement strategy applied in this study can be considered as a combination of steady-state and elitist replacement methods. It helps to keep the best solutions for older generations and to maintain the population diversity to avoid premature convergence. Assume that $\boldsymbol{P}^-$ is the population of the last generation, $\boldsymbol{P}^+$ is the population of the selected parents from $\boldsymbol{P}^-$, and $\boldsymbol{P}^{++}$ is the population of the reproduced offsprings using $\boldsymbol{P}^+$. Accordingly, q^- demonstrates the least fitness value among the best third quartile of individuals in $\boldsymbol{P}^-$. Therefore, the new population $\boldsymbol{P}$ starts forming by the fittest individuals of $\boldsymbol{P}^-$ (elites). Among the elite individuals with similar fitness, those who are formed by matches coming from more distinct regions of the overlapping rectangle have priority in replacement. This way, the chance of ending with the local optima is highly reduced. In fact, this factor is used to distinguish models with similar fitness values. The rest of the places at $\boldsymbol{P}$ are occupied by the following replacement condition.

$$\boldsymbol{P}_i = \begin{cases} \boldsymbol{P}_i^{++} & \text{if fitness}(\boldsymbol{P}_i^{++}) > q^- \\ \boldsymbol{P}_i^{+} & \text{if fitness}(\boldsymbol{P}_i^{++}) < q^- \end{cases} \tag{9}$$

Equation (9) implies that an offspring whose quality is worse than 75% of the previous solutions cannot replace its parents. The genetic algorithm iterates the procedures mentioned above until there is no improvement in the average of the elites' fitness values during a specified number of generations.

4.4.7 Inlier classification

Once the genetic algorithm terminates, the inlier set of minimum cardinality, $\hat{\boldsymbol{I}}$, is found out, and the accurate fundamental matrix, $\hat{F}$, is re-estimated using these points using the following equation by performing iterative least squares adjustment,

$$\begin{aligned} &\boldsymbol{U}(\boldsymbol{\theta},\boldsymbol{l}) = \boldsymbol{u}_i'^T \tilde{F}\ \boldsymbol{u_i} = 0 \qquad\qquad , i \in \hat{\boldsymbol{I}} \\ &\text{where:} \\ &\tilde{\boldsymbol{u}}_{\boldsymbol{i}}' = T'\boldsymbol{x}_{\boldsymbol{i}}',\ \boldsymbol{u_i} = T\boldsymbol{x_i} \\ &\boldsymbol{\theta} = [\tilde{F}_{11} \quad \tilde{F}_{12} \quad \tilde{F}_{13} \quad \tilde{F}_{21} \quad \tilde{F}_{22} \quad \tilde{F}_{23} \quad \tilde{F}_{31} \quad \tilde{F}_{32}] \\ &\boldsymbol{l} = (\boldsymbol{u}_i'^T, \boldsymbol{u}_i^T) \quad , i \in \hat{\boldsymbol{I}} \end{aligned} \tag{10}$$

where T and T' are Hartley normalizing transformations.

Now, inliers can be distinguished as the matches that their residuals from the final fundamental matrix are less than a given threshold. The important issue would be determining this threshold. Standard RANSAC algorithms determine this quantity using maximum likelihood estimation based on the median of the residuals associated with the best tentative model. In this paper, the uncertainty of the final fundamental matrix is used to calculate an adaptive threshold as follows.

The covariance matrix of the estimated parameters can be derived using the covariance law as in Equation (11), where $A = \dfrac{\partial \boldsymbol{U}}{\partial \boldsymbol{\theta}}$, $B = \dfrac{\partial \boldsymbol{U}}{\partial \boldsymbol{l}}$ and $\hat{\boldsymbol{v}}$ is the vector of estimated residuals.

$$\hat{\Sigma}_{\hat{\boldsymbol{\theta}}} = \frac{\hat{\boldsymbol{v}}^{\mathrm{T}}\hat{\boldsymbol{v}}}{n^* - 7}(A^T(BB^T)^{-1}A)^{-1} \tag{11}$$

From Equation (10), $\hat{F} = T'^T \tilde{F} T \triangleq G(\hat{\boldsymbol{\theta}})$. We can determine the uncertainty of the estimated fundamental matrix using the rules of error propagation as follows.

$$\hat{\Sigma}_{\hat{F}} = \left(\frac{\partial G}{\partial \hat{\boldsymbol{\theta}}}\right)\hat{\Sigma}_{\hat{\boldsymbol{\theta}}}\left(\frac{\partial G}{\partial \hat{\boldsymbol{\theta}}}\right)^T \tag{12}$$

For each match $\boldsymbol{x_i} \leftrightarrow \boldsymbol{x_i'}$ belonging to the inlier set of minimum cardinality, the Sampson distance d_{i} and its variance $\sigma_{d_i}^2$ can be calculated via Equation (13).

$$\sigma_{di}^2 = \left(\frac{\partial d_i}{\partial \boldsymbol{y}}\right) \Sigma_{\boldsymbol{y}} \left(\frac{\partial d_i}{\partial \boldsymbol{y}}\right)^{\mathrm{T}} \quad \forall\ i \in \hat{\boldsymbol{I}}$$

where:

$$d_i = (\boldsymbol{x_i'}^{\mathrm{T}} \hat{F} \boldsymbol{x_i})^2 / (l_1^2 + l_2^2 + l_1'^2 + l_2'^2) \tag{13}$$

$$\boldsymbol{y} = [\hat{F}_{11}, ..., \hat{F}_{32}, \hat{F}_{33}, \boldsymbol{x_i}^{\mathrm{T}},\ \boldsymbol{x_i'}^{\mathrm{T}}]^{\mathrm{T}}$$

$$\Sigma_{\boldsymbol{y}} = \begin{bmatrix} \hat{\Sigma}_{\hat{F}} & 0_{9\times 4} \\ 0_{4\times 9} & \sigma_{\max} I_4 \end{bmatrix}$$

The average of these distances, $\mu_d^* = \sum d_i / n^*$, and their standard deviation, $\sigma_d^* = \sqrt{\sum \sigma_{d_i}^2} / n^*$, can represent the distribution of the residuals for inliers. Considering Tchebyscheff's inequality, at least 95% of the population is within 4.47 times the standard deviation from the mean, no matter what kind of probability distribution they are following. Therefore, every match k is an outlier by the confidence of 95% if its residual, d_k, is greater than $\mu_d^* + 4.47\sigma_d^*$.

4.5 Experimental Results and Discussions

To demonstrate the efficiency of our algorithm and its individual components, we performed several experiments on simulated and real data. The *variables* that are tested with these experiments include: i) the performance of our sampling scheme, ii) the accuracy of our adaptive thresholding method for inlier classification, iii) the effect of GA population size on the performance of the algorithm, iv) the performance of the overall algorithm under different levels of noise and outliers. Table 4.6.1 summarizes the *criteria* used to assess these variables.

Table 4.6.1. Criteria for performance assessment

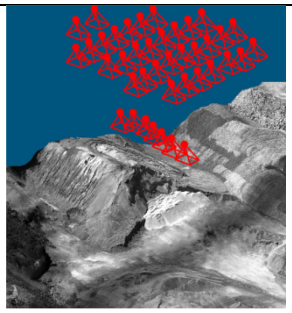
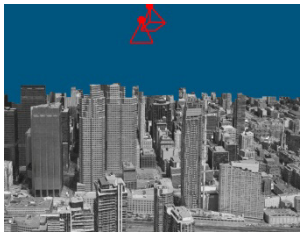
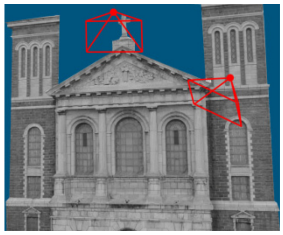
Symbol	Description
Itr	number of iterations before the termination of robust estimation
N_{model}	number of models hypothesized before termination (for most algorithms, equal to the number of sample sets drawn before termination)
VPM	number of verifications per model
ρ	inlier ratio: percentage of the detected inliers from the total number of putative matches
μ_d	precision of estimation: average of squared Sampson residuals over the detected inliers that shows how well the estimated fundamental matrix fits the detected inliers
α	accuracy of inlier classification: percentage of correctly identified outliers and inliers among all the matches (applicable when ground-truth is available)
TPR	sensitivity or true positive rate: percentage of correctly identified inliers (applicable when ground-truth is available)
TNR	specificity or true negative rate: percentage of correctly identified outliers (applicable when ground-truth is available)
μ_d.*CP*	accuracy of estimation: average of squared Sampson residuals on control points or true inliers that shows how well the estimated model fits the real inliers (applicable when ground-truth is available)
D_F	difference between the estimated fundamental matrix and the true one measured using a method described by Zhang (1998) (applicable when ground-truth is available)

The experimental results obtained from the proposed technique are compared with those of the following representative state-of-the-art techniques: RANSAC, MSAC, MLESAC, RRANSAC with SPRT test, LO-RANSAC (Lebeda et al., 2012), PROSAC, StaRSAC, Cov-RANSAC, Multi-GS RANSAC and least median of squares (LMedS).

4.5.1 Experiments on synthetic data

Several synthetic datasets were used in order to evaluate the performance of our algorithm. Using the synthetic data allowed us to control the imaging geometry, the fraction of outliers and the image noise. Besides, it let us assess the accuracy of inlier-detection and model estimation in comparison with the ground truth. Instead of creating random correspondences without having any specific geometric or physical form, real 3D point clouds were used to generate synthetic images. The synthetic datasets are described in Table 4.6.2

Table 4.6.2. Description of synthetic datasets

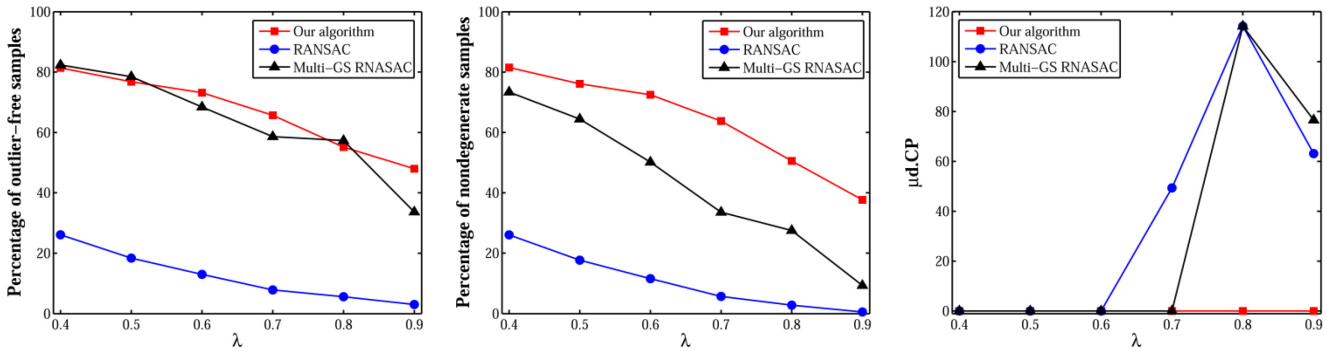
Dataset Label	Figure (3D scene and Cameras Orientation)	Description
Multi-view		– 40 synthetic images – averagely 3000 correspondences per stereo pair – focal length: 4730 pixels, sensor size: 4870×3250 pixels – Gaussian noise: from 0 to 2 pixels – outlier ratio: from 20% to 80% – outlier: gross error from both normal and uniform probability distributions ranging from 10 to 30 pixels
Urban[1]		– 2500 correspondences – focal length: 8889 pixels, sensor size: 5750×3750 – Gaussian noise: from 0 to 4 pixels – outlier ratio: from 0% to 80% – outliers: from 10 to 30 pixels
Church		– 800 correspondences – focal length: 1000 pixels, sensor size: 3000×2000 – Gaussian noise: from 0 to 2 pixels – outlier ratio: from 20% to 70% – outliers: from 10 to 40 pixels
Table		– focal length: 3500 pixels, sensor size: 1940×1460 pixels – Gaussian noise= from 0 to 2 pixels – correspondences are either on the monitor plane or other objects on the table – λ is the number of matches located on the monitor plane divided by the total number of matches – six instances of the data created by varying λ from 0.4

[1] The urban 3D point cloud from which the images are synthesized belong to ISPRS benchmark datasets from Toronto area.

Dataset Label	Figure (3D scene and Cameras Orientation)	Description
		to 0.9 – 258 correspondences on the monitor plane at each instance of data – a total of 100 random outliers at each instance of data

4.5.1.1 Performance of the guided sampling technique

In order to verify the efficiency of our sampling algorithm to avoid planar degeneracy, the Table dataset was used. In each instance of this data, 258 synthesized correspondences were located on the monitor plane (as the dominant plane of the scene), and the number of matches from other objects varied to get the ratio λ. For instance, at λ=0.4, the total number of correspondences was 645, from which 258 correspondences located on the monitor plane. For any instance of the dataset at different ratios λ, we limited the total number of hypotheses to 1000 and compared the results of our method with those of RANSAC and Multi-GS RANSAC. That is each algorithm was stopped when exactly 1000 sample sets were generated. This limited number of samples is approximately 6 times more than the theoretical number of sample sets for achieving 95% probability of picking up at least one outlier-free random sample set from the data containing 40% outliers, which is the maximum outlier ratio in the dataset instances. The performance criteria used for this experiment were i) the percentage of outlier-free[1] sample sets among the 1000 sample sets, ii) the percentage of non-degenerate[2] and outlier-free sample sets among the 1000 sample sets that we denote as non-degenerate sample sets for simplicity, and iii) estimation accuracy (*μd.CP*). The median of the results obtained after five trials is represented in Figure 4.6.1. The proposed algorithm drew up to 75 times more outlier-free sample sets in comparison with RANSAC within a limited budget of 1000 sample sets. The percentages of outlier-free samples for our algorithm and Multi-GS RANSAC were very close. This showed that Multi-GS sampling strategy performed well in the absence of degeneracy. However, Multi-GS sampling failed to draw non-degenerate sample sets as the ratio λ increased. For ratios higher than 0.7, both RANSAC and Multi-GS RANSAC failed to estimate the model correctly (Figure 4.6.3c), while the proposed algorithm estimated the model robustly even in the presence of serious degeneracy (λ=0.9).

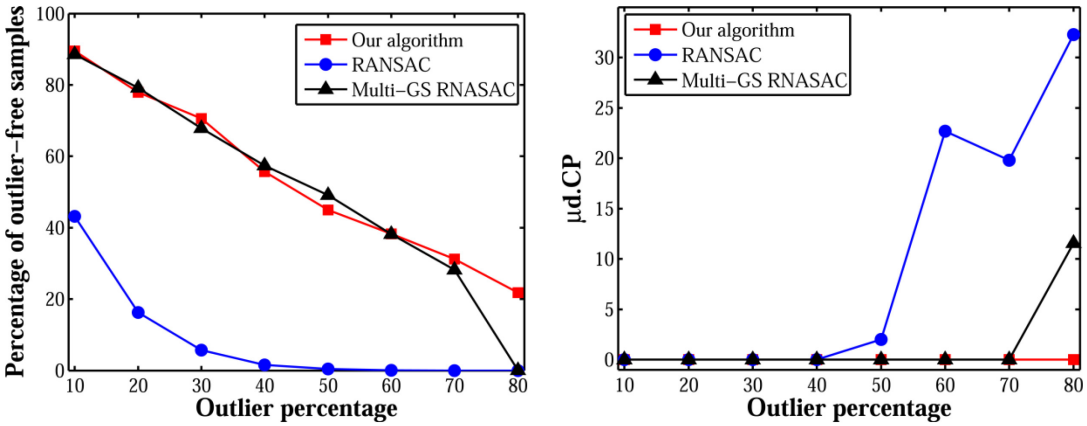


[1] An outlier-free sample set is a set of matches where all the matches are inlier.

[2] A degenerate sample set contains more than five points from the dominant plane (in this example, the monitor plane).

Figure 4.6.1. Performance of sampling methods on Table dataset as the ratio λ increases (a) percentage of outlier-free sample sets, (b) percentage of non-degenerate sample sets, (c) estimation accuracy.

A similar test was performed to assess the performance of the sampling algorithm under various ratios of outliers. To this end, the Church dataset was used. The limited sampling budget was set to 1000 and 5000 sample sets for outlier ratios less than or equal to 50% and higher than 50%, respectively. The percentage of outlier-free sample sets among the budget sample sets (either 1000 or 5000) and the estimation accuracy (*μd.CP*) are presented in Figure 4.6.2.



a b

Figure 4.6.2. Sampling methods on Church dataset as the outlier ratio increases (a) percentage of outlier-free sample sets, (b) estimation accuracy.

For most of the outlier ratios, the outlier-free sampling rates of the proposed algorithm and those of Multi-GS were very close. RANSAC generally drew quite less number of outlier-free sample sets in comparison with the other two methods. For outlier ratios higher than 50%, RANSAC completely failed to detect outlier-free sample sets and to estimate the model correctly. For outlier ratio of 80%, only the proposed algorithm kept good performance by drawing at least 22% outlier-free sample sets, given the limited sampling budget. As a conclusion, the sampling strategy based on the spatial distribution of matches along with the evolutionary search not only increased the speed of reaching an outlier-free sample set but also decreased the probability of ending up with a degenerate solution.

4.5.1.2 Performance of inlier classification with adaptive thresholding

In order to verify the performance of the proposed thresholding method for inlier classification (Section 4.4.7), a stereo pair from Multi-view dataset at baseline of 20 meters was used. To eliminate the effect of other components of the algorithm, such as sampling and objective function, no outlier was introduced to the images; i.e. the dataset was outlier-free. The results from our method were compared with those of the median-based and covariance-based algorithms. To this end, each algorithm was applied 500 times. In any trial, a minimal sample set of 8 points was randomly drawn, and the fundamental matrix was estimated with the normalized 8-point algorithm. For our algorithm, the technique of Section 4.4.7 was

applied to determine the inlier threshold. For the median-based algorithm, the robust standard deviation of residuals was determined as $\sigma = 1.4826(1+5/(n-8))\sqrt{median_i(r_i)}$. As the dataset was outlier-free, n was equal to the total number of matches, and r_i for $i=[1:n]$ was the residual of any correspondence. Then, the inlier threshold was calculated as 1.96σ . For the covariance-based algorithm, which is the core component of Cov-RANSAC, first, the uncertainty of the estimated model was used to narrow down the total set of matches to the potential inliers. Then the median-based algorithm was applied to the potential inliers for determining the inlier threshold.

In order to evaluate the performance of these algorithms, the fraction of runs (from 500 runs), in which a correspondence was classified as inlier was calculated, namely the inlier probability of that match. The inlier probabilities (sorted in ascending order) are shown in Figure 4.6.3. Knowing that the dataset was outlier-free, the inlier probability for all the points was ideally one. Our thresholding algorithm resulted in the most stable and robust solutions. That is, for 92% of the points, the inlier probability was higher than 0.9. For the median-based algorithm, this percentage was only 15%. The covariance-based algorithm had poor performance in comparison with both other methods. Our algorithm resulted in inlier ratios higher than 90% at more than 88% of the runs. However, the median-based and covariance-based algorithms yielded inlier ratios higher than 90% at only 32% and 15% of the runs, respectively.

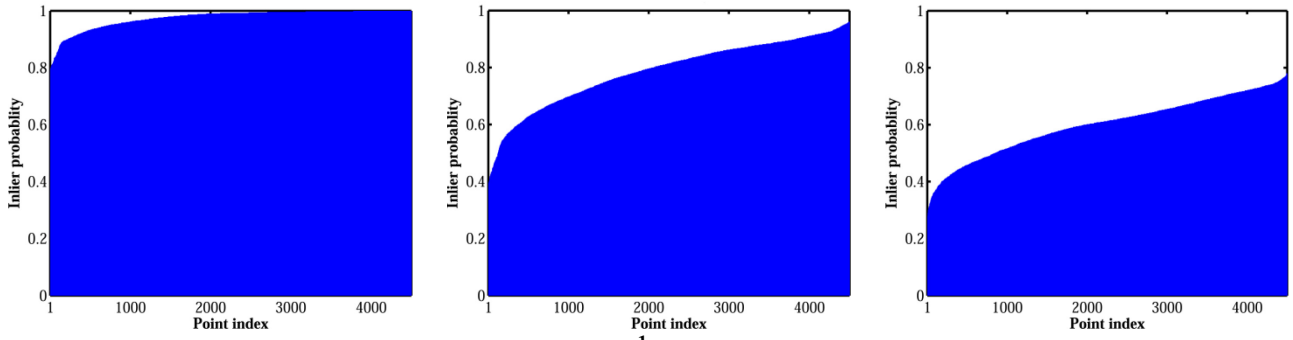


Figure 4.6.3. Inlier probability of correspondences obtained using (a) our adaptive thresholding method, (b) median-based algorithm, (c) covariance-based algorithm.

To investigate the reason behind the performance of each algorithm, the threshold values determined at each run are illustrated versus the run index in Figure 4.6.4. The maximum and median of the residuals from the true fundamental matrix are shown too. Using the median-based algorithm, the threshold value based on the true residuals was calculated and is illustrated in Figure 4.6.4. The true fundamental matrix was directly calculated from the orientation values used for synthesizing these images.

The true median-based threshold was slightly higher than the true maximum residual, and it would be an ideal choice of threshold only if the fundamental matrix were perfect. However, in reality, the fundamental matrix was calculated from a minimal sample set of eight correspondences, which had different values of noise and did not necessarily have an ideal spatial configuration either. Therefore, the uncertainty of the fundamental matrix and the noise value of each point could affect the threshold value.

Although the covariance-based method tried to consider this effect, it underestimated the potential inliers. The main reason was that the uncertainty of the fundamental matrix was estimated only from the minimal sample set. As shown in Figure 4.6.4, our algorithm varied the threshold adaptively. Frequently, the threshold value was slightly higher than the ideal median-based threshold. This was reasonable since the calculated fundamental matrix, although estimated from an outlier-free sample set, was not necessarily perfect. However, for the other two methods, the threshold values were approximately around the true median-based threshold, which could be a good threshold only if the fundamental matrix were as accurate as the true one.

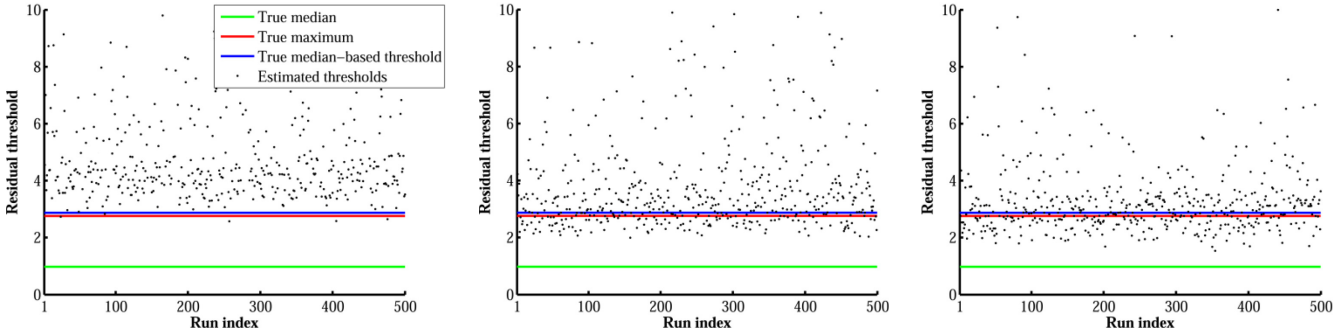


Figure 4.6.4. Inlier thresholds determined by (a) our adaptive thresholding method, (b) median-based algorithm, (c) covariance-based algorithm.

4.5.1.3 Performance under various outlier ratios

In order to assess the performance of the overall algorithm under various ratios of outliers, the Multi-view data set was used. In the following experiments, the stall generation of GA (number of generations during which the best fitness should not change before termination) was set to 60, and the maximum number of iterations was set to 1000. The upper bound to the standard deviation of image noise (σ_{max}) was set to 3 pixels. The GA population size was set to 27. The parameter n^* was set to $n/10$ in all our experiments. All the algorithms were implemented in MATLAB directly without using the optimization toolbox. The average of the results over all the stereo pairs versus the outlier ratios are presented in Figure 4.6.5.

On the downside, the number of models hypothesized by our algorithm was relatively higher than other methods, especially for low percentages of outliers. This was caused by the fact that GA terminated when no improvement was achieved after a certain number of iterations. This was not a suitable termination criterion when the ratio of inliers in the dataset was very high, since, at any iteration, it was highly probable to encounter another outlier-free sample set from which a slightly better model could be estimated. Therefore, the improvement stopped after a higher number of iterations. However, this was advantageous in high outlier ratios. For instance, to reach 95% accuracy in inlier detection from a dataset with 70% outliers, at least 45658 sample sets must be drawn in random sample consensus. However, the proposed algorithm reached 95% accuracy by drawing only 2100 sample sets. Similarly, our algorithm achieved 78% accuracy over the 80% contaminated dataset with only 1440 hypotheses,

while 591455 random sample sets would be theoretically required to reach that accuracy. In addition, for all the outlier ratios, the number of GA iterations was lower than other algorithms (average *Itr* of 93). Since the genetic algorithm can be parallelized easily, we could benefit from the low number of iterations and parallelize all the sampling procedures within any iteration. This could eventually result in higher computational efficiency in comparison with other methods.

In terms of the number of evaluations per models, the proposed algorithm behaved quite like other variants of RANSAC, except for RRANSAC. Since RRANSAC benefited from the SPRT test, each model was not necessarily evaluated against all the points. Therefore, for RRNSAC, the average number of verifications per model was noticeably lower than other methods. However, the required number of hypotheses was noticeably more than other methods, since good hypotheses might have been ignored due to randomized verification.

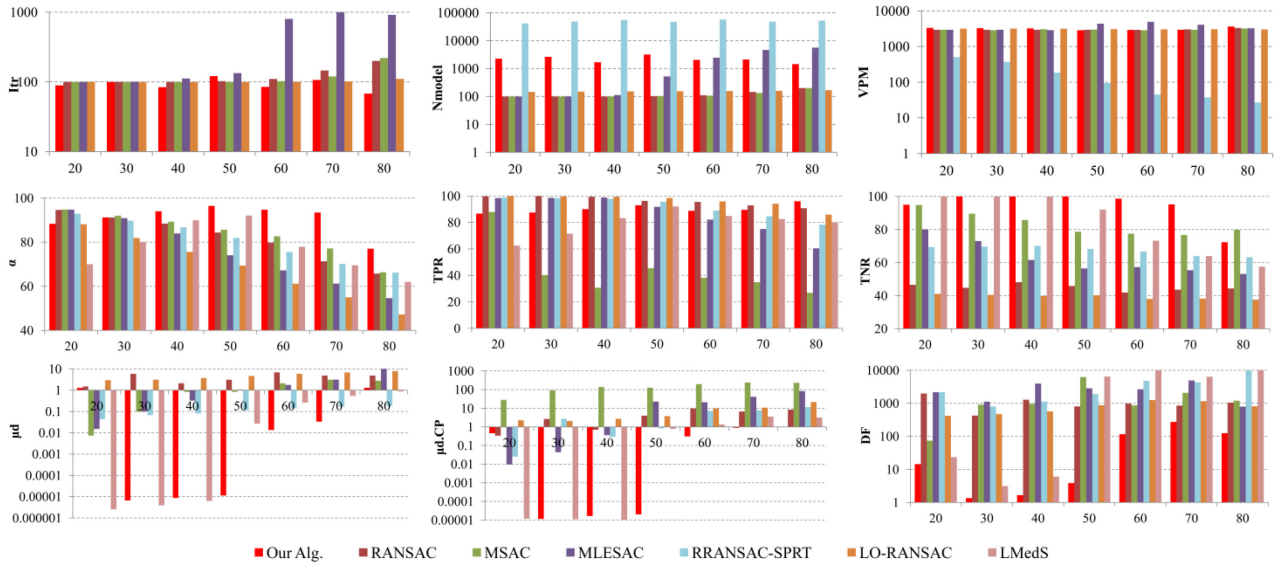


Figure 4.6.5. Performance of different algorithms under various percentages of outliers for Multi-view dataset. In the graphs, the x-axis represents the percentage of synthetic outliers in the dataset.

From the accuracy point of view, the proposed algorithm was more robust to outliers in comparison with other algorithms, especially when the outlier percentage grew over 40%. On average, the proposed technique achieved 91%±6% accuracy for inlier detection. The considerably high estimation accuracy (average $\mu_d.CP$ of 0.376) confirmed this as well. The improvement achieved by the proposed algorithm was also obvious in terms of true negative rate (average TNR of 94%), which showed the efficiency of the algorithm to distinguish outliers from inliers.

4.5.1.4 Performance stability

Another variable that should be tested under various outlier ratios is the stability of the results obtained by running the algorithm multiple times. To this end, the Church dataset was used. The algorithm was

performed 50 times on each instance of the data. The average and standard deviation of the results obtained from these runs are shown in Figure 4.6.6.

For outlier ratios ranging from 10% to 70%, the variations of accuracy was not considerable (α=99%±2%). However, at 80% outlier rate, the average accuracy decreased to 92%±8%. Although this was a noticeable change in the performance of the algorithm, the median of the accuracies was yet reasonably high (95%). Regarding the number of models hypothesized before termination, high range of changes was observed at lower outlier ratios. As explained earlier, this happened due to the stopping criterion in GA. However, this was itself an advantage for large outlier ratios (more than 50% outliers).

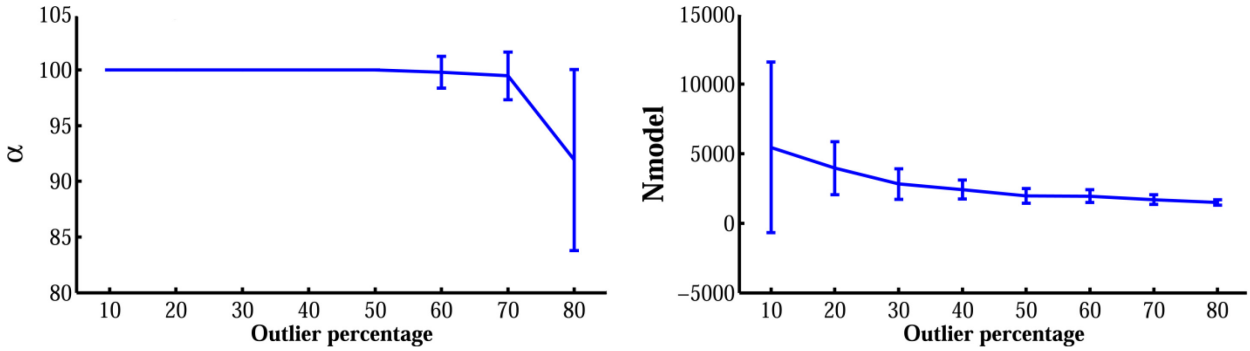


Figure 4.6.6. Performance of the proposed algorithm from the stability point of view.

4.5.1.5 Performance under various noise levels

The Urban dataset was used to evaluate the robustness of the proposed algorithm against noisy image observations. To this end, the effects of image noise along with varying percentages of outliers were tested using 81 instances of Urban data set. The results are illustrated in Figure 4.6.7.

From the accuracy point of view, the accuracy of estimation (represented by $\mu_{d.CP}$) did not decrease by increasing the noise level. In terms of accuracy of inlier detection, the true positive rate was not affected by the noise. However, the true negative rate decreased slightly by increasing the noise level. Since the outliers were synthesized by adding gross errors as low as 10 pixels to correspondences, distinguishing real outliers from noisy inliers at higher noise levels became a more difficult task. In terms of convergence speed, increasing the magnitude of image noise decreased the number of iterations. The total number of iterations required to find the final solution changed only from 80 to 250 as the ratio of outliers increases to 80% and the level of noise increases to 4 pixels. This shows the high computational efficiency of the overall algorithm in the presence of high ratios of outliers and noise.

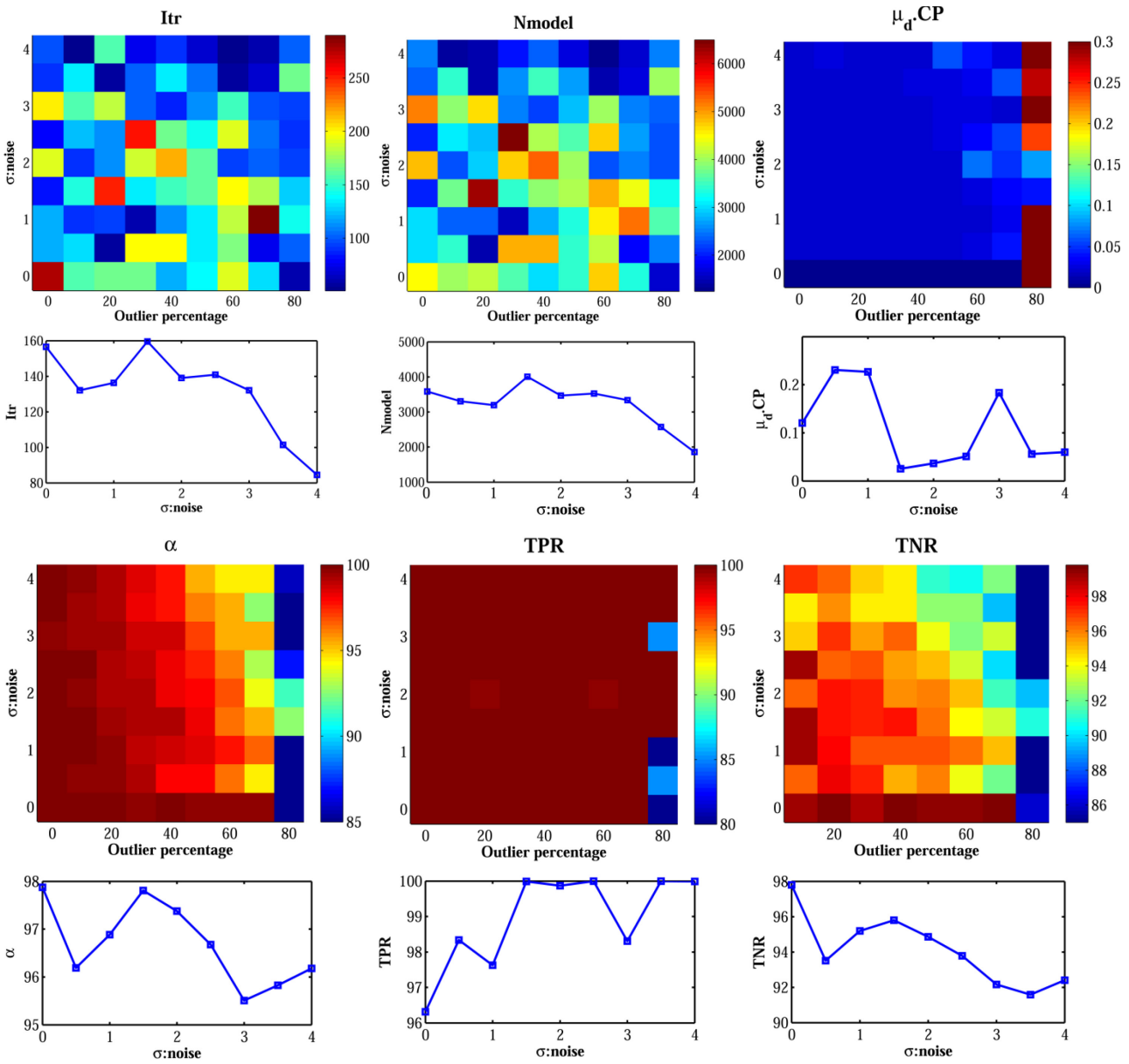


Figure 4.6.7. Performance of the proposed algorithm with noisy images. The graph at the bottom of each surface-plot represents the average of respective performance criterion versus the amount of noise.

4.5.1.6 Effect of the GA population size

In order to evaluate the effect of GA population size on the performance of the algorithm, Church dataset was applied. The GA population size was varied from 20 to 70 by steps of five individuals, and the performance of the algorithm was assessed under various outlier ratios. The results obtained from this experiment are illustrated in Figure 4.6.8.

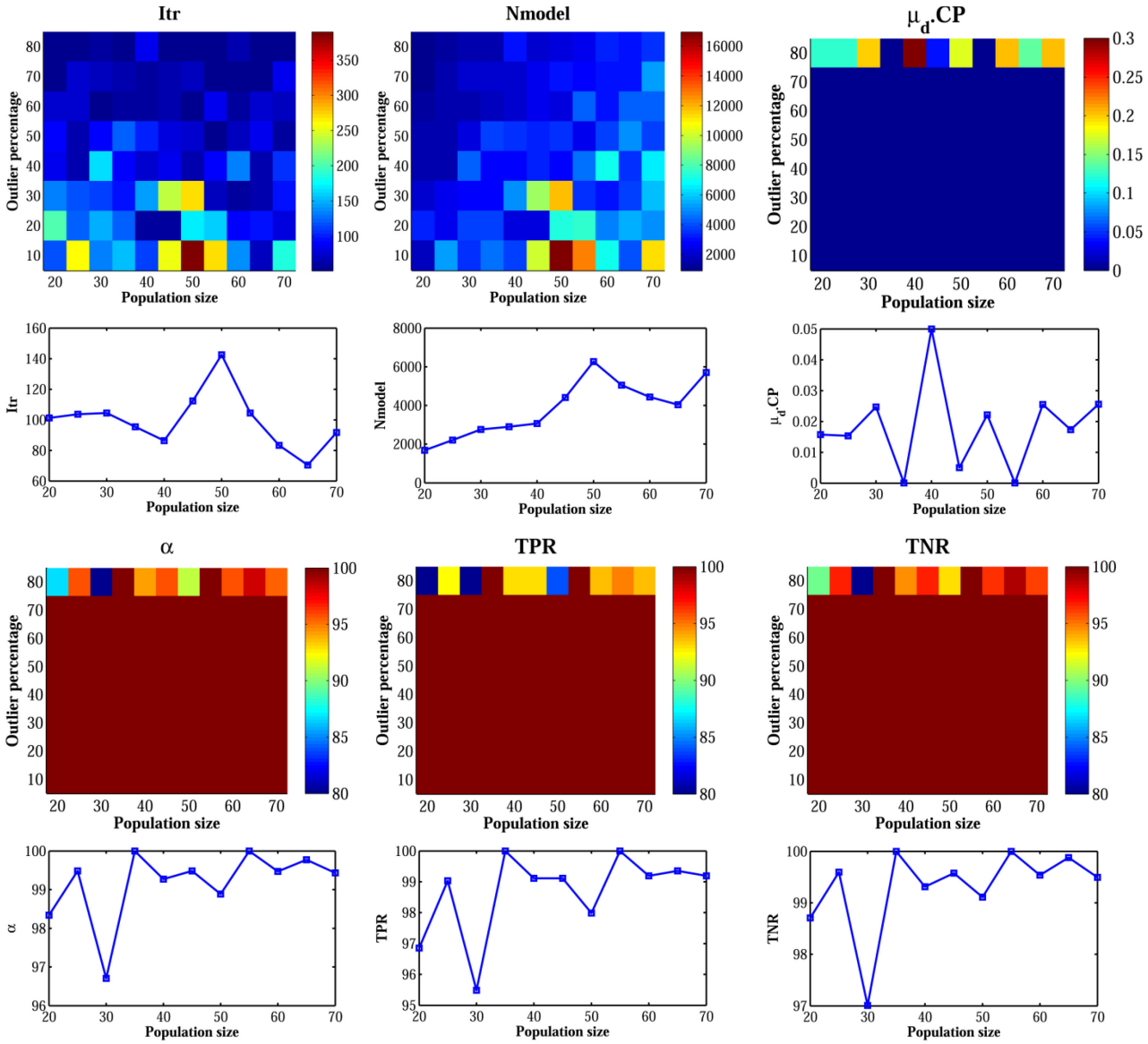


Figure 4.6.8. Performance of the proposed algorithm with varying GA population size. The graph at the bottom of each surface-plot represents the average of the respective performance criterion versus the population size.

There was no considerable correlation between the accuracy and the size of the population. The average accuracy of inlier-detection (α) with different population sizes was 99.1% with standard deviation of only 0.9%. The accuracy of estimation ($\mu_d.CP$) also remained below 0.05, which confirmed the stable accuracy of model estimation. The number of iterations before termination (*Itr*) seemed to be more dependent on the outlier percentage rather than the population size. The average number of iterations at outlier ratios less than 50% was 122, while it was only 62 at higher outlier ratios. The number of generated hypotheses (*Nmodel*) increased by either decreasing the outlier ratio or increasing the GA population size since it depended both on the population size and number of iterations. Showing that the

accuracy does not have any specific correlation with the population size, the low size of the population is suggested to both avoid unnecessary generation of hypotheses and to keep the number of iterations optimal.

4.5.2 Experiments on real data

The proposed algorithm of robust estimation and inlier detection was compared with other RANSAC-like techniques for 15 real stereo pairs (Table 4.6.3). These examples were chosen to cover various cases, such as close-range and aerial photography, narrow and wide baselines, degenerate configurations, scale variation, multi-platform photography and highly contaminated data. The first nine pairs and their putative matches were gathered from Raguram et al. (2013). For these data, the percentages of true inliers in the dataset (ε) were manually determined in the reference article. For the 10th, the 11th and the 12th stereo pairs, we used signalized targets to provide ground control data when acquiring the images. The reference fundamental matrices were calculated from these control points. The 13th stereo pair belongs to the ISPRS data sets for urban classification[1]. The 14th and the last stereo pairs belong to the ISPRS benchmark for multiplatform photogrammetry[2] (Nex et al., 2015). For these last three pairs, the exterior and interior orientation parameters provided by ISPRS were used as reference values for evaluation. For the last seven stereo pairs, SIFT key points were detected and matched using VlFeat[3] feature-based matching library. For the 4th and the 14th stereo pairs, we could not apply Multi-GS, since the low inlier ratio and a high number of correspondences made that algorithm too slow to be executable.

It can be noticed from the results that our algorithm generally yielded solutions that were compatible with the ground-truth in terms of either inlier-detection accuracy or estimation accuracy. The main drawback of the proposed method, compared with other methods, was the higher number of models hypothesized (N_{model}) for high inlier ratios. This was due to both the termination criterion in GA and the fact that more than one model was hypothesized at any iteration. However, it can be still noticed that the total number of iterations (*Itr*) for the proposed algorithm was either close to or noticeably lower than other methods. Therefore, in terms of computational time, this drawback could be compensated by parallelizing the iterations. Besides, the algorithm benefited from the fast linear estimation of fundamental matrix based on partial pivoting, which was computationally a light operation. For the 10th and the 11th images, a dominant plane existed in the scene. Therefore, most of the algorithms ended up to a degenerate solution. However, the degeneracy could be avoided quite efficiently by our algorithm. This was mainly due to the guided sampling based on spatial distribution of matches. Finally, it was noticed that the algorithm had good performance in challenging cases such as multi-platform photography. In general, the accuracy of the results was most close to that of ProSAC and RRANSAC, however, with much higher computational efficiency.

[1] www2.isprs.org/commissions/comm3/wg4/detection-and-reconstruction.html

[2] http://www2.isprs.org/commissions/comm1/icwg15b/benchmark_main.html

[3] www.vlfeat.org

Table 4.6.3. Performance of the proposed algorithm and other techniques on real data

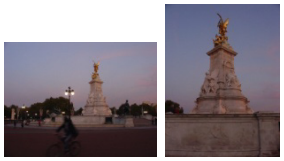
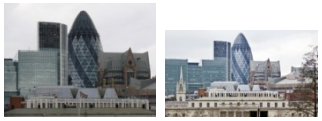
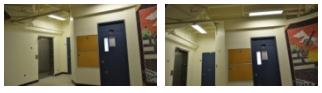
Image pairs		Param.	Our Alg.	RANSAC	MSAC	RRANSAC -SPRT	LO-RNASAC	Cov-RANSAC	ProSAC	StaRSAC	MultiGS RANSAC
1	ε=48% , n=3154* 	*Itr*	103	814	743	50001	1984	502	401	76295	824
		Nmodel	2579	814	743	50000	2093	501	406	76295	824
		VPM	3127	3154	3154	15	3169	16	3115	3154	3154
		ρ	47.0	43.5	43.8	37.8	45.0	0.3	54.2	50.3	49.6
		μd	0.192	0.160	0.116	0.143	0.262	972.343	6.982	0.824	0.824
2	ε=57% , n=575 	*Itr*	116	304	219	50001	824	503	187	38815	111
		Nmodel	2801	304	219	50000	922	502	190	38815	111
		VPM	553	575	575	6	581	20	566	575	575
		ρ	54.6	54.6	52.9	54.3	56.2	1.9	59.7	60.5	78.5
		μd	0.235	0.122	0.116	0.116	0.168	9.133	1.764	3.148	8.122
3	ε:38% , n=1088 	*Itr*	82	2823	3201	50001	16038	503	1400	160818	2645
		Nmodel	1988	2823	3201	50000	16092	502	1409	160818	2645
		VPM	1062	1088	1088	9	1088	25	1081	1088	1088
		ρ	34.0	32.0	32.0	31.5	34.2	1.1	47.2	43.2	42.8
		μd	0.163	0.111	0.092	0.120	0.228	24.916	16.625	6.076	28.593
4	ε=22% , n=1516 	*Itr*	203	46329	50000	500001	500000	505	5001	112200	
		Nmodel	4962	46329	50000	500000	500068	504	5010	112200	
		VPM	1482	1516	1516	266	1516	21	1513	1516	
		ρ	21.6	8.0	20.6	5.9	6.9	0.7	33.3	14.9	
		μd	0.163	0.208	0.074	0.137	0.455	139.683	115.022	15.881	
5	ε=38% , n=422 	*Itr*	80	217	1044	50001	18757	503	1177	91638	198
		Nmodel	1819	217	1044	50000	18921	502	1185	91638	198
		VPM	415	422	422	4	422	12	419	422	422
		ρ	31.8	32.7	33.4	24.9	34.8	2.6	47.4	49.5	100.0
		μd	0.199	0.115	0.170	0.306	0.194	154.951	24.888	46.196	4911
6	ε=45% , n=904 	*Itr*	78	964	1100	50001	2547	502	326	52692	1259
		Nmodel	1885	964	1100	50000	2645	501	329	52692	1259
		VPM	888	904	904	11	907	15	896	904	904
		ρ	44.1	42.3	40.5	38.9	43.0	1.4	55.6	59.5	43.7
		μd	0.179	0.102	0.100	0.110	0.250	0.614	18.828	32.472	20.919
7	ε=85% , n=2471 	*Itr*	80	100	100	50001	100	502	101	6211	101
		Nmodel	1988	100	100	50000	143	501	105	6211	101
		VPM	2398	2471	2471	218	2540	423	2377	2471	2471
		ρ	83.9	76.1	75.4	77.1	82.6	17.0	93.0	61.4	93.1
		μd	0.155	0.139	0.144	0.091	0.332	0.048	0.377	0.053	0.373
8	ε=65% , n=1110 	*Itr*	63	100	100	50001	134	502	101	17595	101
		Nmodel	1526	100	100	50000	177	501	105	17595	101
		VPM	1095	1110	1110	35	1135	104	1068	1110	1110
		ρ	66.4	55.0	57.5	62.3	66.4	9.2	73.1	58.5	74.4
		μd	0.252	0.303	0.219	0.169	0.369	0.081	0.890	0.187	0.605
9	ε=60% , n=962 	*Itr*	178	100	100	50001	144	502	101	14122	101
		Nmodel	4089	100	100	50000	209	501	103	14122	101
		VPM	757	962	962	112	990	101	943	962	962
		ρ	60.7	65.5	56.3	61.3	62.2	10.2	71.5	72.6	70.5
		μd	0.209	0.613	0.136	0.141	0.254	0.099	27.357	4.009	1.103
10	ε=N/A , n=1497 	*Itr*	105	100	100	50001	100	502	101	6028	101
		Nmodel	2520	100	100	50000	176	501	106	6028	101
		VPM	1433	1497	1497	68	1557	464	1547	1497	1497
		ρ	89.9	96.8	81.0	93.8	98.1	17.3	90.3	97.2	90.5

Image pairs		Param.	Our Alg.	RANSAC	MSAC	RRANSAC -SPRT	LO-RNASAC	Cov-RANSAC	ProSAC	StaRSAC	MultiGS RANSAC
		μd	0.122	5.911	1.908	4.638	26.657	0.053	1.570	248.279	0.621
		D_F	4.0	147.2	234.5	156.1	235.3	2433.3	34.2	122.3	9036.2
		$\mu d.CP$	0.335	543.293	750.871	371.097	293.029	9052.456	23.365	112.079	30254.320
11	ε=N/A , n=4355	*Itr*	120	100	100	22595	100	502	101	6000	101
		Nmodel	2746	100	100	22594	132	501	107	6000	101
		VPM	3618	4355	4355	4	4454	1646	4111	4355	4355
		ρ	94.7	84.7	89.1	94.7	97.4	37.7	97.0	97.7	96.9
		μd	0.071	0.093	0.077	0.075	0.744	0.041	0.117	0.435	0.117
		D_F	2.5	284.9	371.3	117.7	76.3	733.4	55.4	93.0	96.9
		$\mu d.CP$	0.737	3138.844	5820.395	517.842	233.113	33732.238	109.322	298.400	214.419
12	ε=N/A , n=1647	*Itr*	59	100	100	50001	100	502	109	27292	101
		Nmodel	1448	100	100	50000	187	501	118	27292	101
		VPM	1571	1647	1647	3	1717	67	1521	1647	1647
		ρ	60.7	88.4	73.3	35.3	89.9	3.8	75.6	92.0	83.5
		μd	0.309	3.447	1.986	0.224	2.821	0.061	3.974	7.403	2.865
		D_F	13.7	80.6	74.2	172.3	70.1	806.2	42.4	96.1	94.6
		$\mu d.CP$	1.356	2.707	2.774	11.226	1.218	4459.101	1.975	4.675	4.134
13	ε=N/A , n=1213	*Itr*	74	100	100	50001	100	502	101	26343	101
		Nmodel	1586	100	100	50000	176	501	104	26343	101
		VPM	1119	1213	1213	85	1261	77	1178	1213	1213
		ρ	37.3	44.2	45.5	50.2	99.3	21.9	93.0	98.0	73.6
		μd	0.258	0.255	0.252	0.267	9.747	0.085	3.043	6.251	3.065
		D_F	10.8	38.2	51.5	22.4	119.6	150.5	14.1	26.0	131.4
14	ε=N/A , n=582	*Itr*	207	50000	5035	50001	8965	507	3374	310000	
		Nmodel	5088	50000	5035	50000	9074	506	3388	310000	
		VPM	229	582	582	5	583	55	580	582	
		ρ	24.9	5.5	2.7	21.6	36.8	8.4	42.4	38.0	
		μd	4.001	6377.040	6.921	0.210	5.953	0.106	201.988	9.657	
		D_F	159.3	1750.5	367.6	188.4	509.0	179.3	1476.6	556.0	
15	ε=N/A , n=3487	*Itr*	346	100	100	50001	100	502	101	7329	101
		Nmodel	8699	100	100	50000	132	501	104	7329	101
		VPM	3334	3487	3487	152	3566	126	3386	3487	3487
		ρ	81.4	89.3	87.6	81.9	90.0	3.4	90.3	90.9	100.0
		μd	0.188	0.382	0.311	0.138	2.536	1.742	0.738	3.767	76190.183
		D_F	12.5	12.4	12.3	12.4	76.1	1555.6	14.4	49.3	2343.3

*ε is the inlier ratio, and n is the total number of correspondences.

4.6 Conclusion

In this study, we proposed a modified version of the integer-coded genetic algorithm for the problem of accurate epipolar-geometry estimation from putative matches, followed by an adaptive thresholding algorithm for inlier classification. The proposed algorithm can be considered as a solution to resolve some of the drawbacks involved in conventional robust estimators, specifically RANSAC-like methods. Based on the experiments, the proposed approach showed robustness to high percentages of outliers, to degeneracy and to image noise. These characteristics are mostly due to the guided sampling strategy, the evolutionary search and the method for inlier classification based on adaptive thresholding. In general, the algorithm was able to detect the inliers by more than 80% accuracy, which is a remarkable success for large data sets. In addition, the computational expenses of the algorithm were not increasing with the ratio of outliers or the magnitude of image noise. As a negative characteristic, the number of models to

be hypothesized before reaching the termination criterion was generally higher than other methods for datasets containing less than 50% outliers. However, this adverse effect can be compensated by reducing the computation time via parallelizing the procedures inside any iteration of GA. Besides, the efficiency of the proposed algorithm in terms of speed and accuracy was noticeably higher than other methods for datasets with a high ratio of outliers (with more than 50% outliers).

Acknowledgement

This study is supported in part by grants from the: Centre de Géomatique du Québec (www.cgq.qc.ca), Fonds de Recherche Québécois sur la Nature et les Technologies (www.frqnt.gouv.qc.ca), and Natural Sciences and Engineering Research Council of Canada (www.nserc-crsng.gc.ca). The authors would like to acknowledge and thank Professor Dr. Wolfgang Förstner for his great ideas to improve the article and his generous devotion of his time to revise this manuscript. The authors would like to acknowledge the provision of the datasets (the 14th and the 15th pairs in Table 4.6.3) by ISPRS and EuroSDR, released in conjunction with the ISPRS scientific initiative 2014 and 2015, lead by ISPRS ICWG I/Vb as well as the provision of the Downtown Toronto datasets (the 3D point cloud used to simulate Urban data set and the 13th pair in Table 4.6.3) by Optech Inc., First Base Solutions Inc., GeoICT Lab at York University, and ISPRS WG III/4.

4.7 References

Apollonio, F. I., Ballabeni, A., Gaiani, M. and Remondino, F., 2014. Evaluation of feature-based methods for automated network orientation. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Riva del Garda, Italy, Vol. XL, Part 5, pp. 47-54.

Armangué, X. and Salvi, J., 2003. Overall view regarding fundamental matrix estimation. Image and Vision Computing, 21(2), pp. 205-220.

Bab-Hadiashar, A. and Suter, D., 1999. Robust segmentation of visual data using ranked unbiased scale estimate. *Robotica*, 17(06), pp. 649-660.

Brown, M. Z., Burschka, D. and Hager, G. D., 2003. Advances in computational stereo. *IEEE T Pattern Anal*, 25(8), pp. 993-1008.

Capel, D., 2005. An effective bail-out test for RANSAC consensus scoring. British Machine Vision Association Conference, http://www.comp.leeds.ac.uk/bmvc2008/proceedings/2005/papers/224/Capel224.pdf (20 Dec. 2014).

Cheng, C. M. and Lai, S. H., 2009. A consensus sampling technique for fast and robust model fitting. *Pattern Recogn*, 42(7), pp. 1318-1329.

Chesi, G., Garulli, A., Vicino, A. and Cipolla, R., 2002. Estimating the fundamental matrix via constrained least-squares: A convex approach. *IEEE T Pattern Anal*, 24(3), pp. 397-401.

Chin, T. J., Yu, J. and Suter, D., 2012. Accelerated hypothesis generation for multistructure data via preference analysis. *IEEE T Pattern Anal*, 34(4), pp. 625-638.

Choi, J. and Medioni, G., 2009. StaRSAC: Stable random sample consensus for parameter estimation. In: In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Miami, pp. 675-682.

Chum, O., Matas, J. and Kittler, J., 2003. Locally optimized RANSAC. *Lecture Notes in Computer Science*, Vol. 2781, pp. 236-243.

Chum, O. and Matas, J., 2005. Matching with PROSAC-progressive sample consensus. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 220-226.

Chum, O. and Matas, J., 2008. Optimal randomized RANSAC. *IEEE T Pattern Anal*, 30(8), pp. 1472-1482.

Chum, O., Werner, T. and Matas, J., 2005. Two-view geometry estimation unaffected by a dominant plane. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 772-779.

Deep, K., Singh, K. P., Kansal, M. L. and Mohan, C., 2009. A real coded genetic algorithm for solving integer and mixed integer optimization problems. *Appl Math Comput*, 212(2), pp. 505-518.

Elbeltagi, E., Hegazy, T. and Grierson, D., 2005. Comparison among five evolutionary-based optimization algorithms. *Adv Eng Inform*, 19(1), pp. 43-53.

Frahm, J. M. and Pollefeys, M., 2006. RANSAC for quasi-degenerate data (QDEGSAC). In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 453-460.

Golub, G. H. and Van Loan, C. F., 2012. *Matrix Computations, Fourth Edition*. The Johns Hopkins University Press, Baltimore, pp. 493 and 262.

Hartley, R. I., 1997. In defense of the eight-point algorithm. *IEEE T Pattern Anal*, 19(6), pp. 580-593.

Hartley, R. I. and Kahl, F., 2007. Global optimization through searching rotation space and optimal estimation of the essential matrix. In: *Proceeding of IEEE International Conference on Computer Vision*, Rio de Janeiro, pp. 1-8.

Hartley, R. and Zisserman, A., 2003. *Multiple View Geometry in Computer Vision*. Cambridge University Press, UK, pp. 239-259.

Hartmann, J., Klussendorff, J. H. and Maehle, E., 2013. A comparison of feature descriptors for visual SLAM. In: *Proceedings of IEEE European Conference on Mobile Robots*, Barcelona, pp. 56-61.

Hubert, M., Rousseeuw, P. J. and Van Aelst, S., 2008. High-breakdown robust multivariate methods. *Stat Sci*, 23(1), pp. 92-119.

Kekec, T., Yildirim, A. and Unel, M., 2014. A new approach to real-time mosaicing of aerial images. *Robot Auton Syst*, 62(12), pp. 1755-1767.

Lebeda, K., Matas, J. and Chum O., 2012. Fixing the locally optimized RANSAC. In: *Proceedings of British Machine Vision Conference*, Guildford, pp. 1-11.

Li, J. and Allinson, N. M., 2008. A comprehensive review of current local features for computer vision. *Neurocomputing*, 71(10), pp. 1771-1787.

Matas, J. and Chum, O., 2004. Randomized RANSAC with $T_{d,d}$ test. *Image Vision Comput*, 22(10), pp. 837-842.

Matas, J. and Chum, O., 2005. Randomized RANSAC with sequential probability ratio test. In: *IEEE International Conference on Computer Vision*, Beijing, pp. 1727-1732.

Myatt, D. R., Torr, P., Nasuto, S. J., Bishop, J. M. and Craddock, R., 2002. NAPSAC: High noise, high dimensional robust estimation. In: *Proceedings of British Machine Vision Conference*, pp. 458-467.

Nex, F., Gerke, M., Remondino, F., Przybilla H. J., Bäumker, M. and Zurhorst, A., 2015. ISPRS benchmark for multi-platform photogrammetry. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-3/W4, pp.135-142.

Ni, K., Jin, H. and Dellaert, F., 2009. GroupSAC: Efficient consensus in the presence of groupings. In: *IEEE International Conference on Computer Vision*, Kyoto, pp. 2193-2200.

Nistér, D., 2005. Preemptive RANSAC for live structure and motion estimation. *Mach Vision Appl*, 16(5), pp. 321-329.

Raguram, R., Chum, O., Pollefeys, M., Matas, J. and Frahm, J., 2013. USAC: a universal framework for random sample consensus. *IEEE T Pattern Anal*, 35(8), pp. 2022-2038. http://www.cs.unc.edu/~rraguram/usac/USAC-1.0.zip

Raguram, R., Frahm, J. M. and Pollefeys, M., 2009. Exploiting uncertainty in random sample consensus. In: *Proceedings of IEEE Conference on Computer Vision*, Kyoto, pp. 2074-2081.

Rodehorst, V. and Hellwich, O. 2006. Genetic algorithm sample consensus (GASAC)-a parallel strategy for robust parameter estimation. In: *Proceedings of IEEE Computer Vision and Pattern Recognition Workshop*, New York City, NY, USA, pp. 103-103.

Sattler, T., Leibe, B. and Kobbelt, L., 2009. SCRAMSAC: Improving RANSAC's efficiency with a spatial consistency filter. In: *IEEE International Conference on Computer Vision*, Kyoto, pp. 2090 - 2097.

Schmidt, A., 2014. The EKF-based visual SLAM system with relative map orientation measurements. *Lecture Notes in Computer Science*, Vol. 8671, 2014, pp. 570-577.

Shahbazi, M., Sohn, G., Théau, J. and Ménard, P., 2015. Robust sparse matching and motion estimation using genetic algorithms. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Munich, Germany, Vol. XL-3/W2, pp. 197-204.

Torr, P. and Murray, D.W., 1997. The development and comparison of robust methods for estimating the fundamental matrix. *Int J Comput Vision*, 24(3), pp. 271-300.

Torr, P. and Zisserman, A., 1998. Robust computation and parametrization of multiple view relations. In: *Proceedings of IEEE International Conference on Computer Vision*, Bombay, pp. 727-732.

Torr, P. and Zisserman, A., 2000. MLESAC: a new robust estimator with application to estimating image geometry, *Comput Vis Image Und*, 78(1), pp.138-156.

Torr, P., 2002. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *Int J Comput Vision*, 50(1), pp. 35-61.

Wendel, A., Maurer, M., Graber, G., Pock, T. and Bischof, H., 2012. Dense reconstruction on-the-fly. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, pp. 1450-1457.

Wöhler, C., 2013. *3D Computer Vision: Efficient Methods and Applications* (2nd ed.). Springer Science & Business Media, pp. 55-73.

Zhang, Z., 1998. Determining the epipolar geometry and its uncertainty: A review. *Int J Comput Vision*, 27(2), pp. 161-195.

Zheng, Y., Sugimoto, S. and Okutomi, M., 2013. A practical rank-constrained eight-point algorithm for fundamental matrix estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, pp. 1546-1553.

Zhong, H. X., Pang, Y. J. and Feng, Y. P., 2006. A new approach to estimating fundamental matrix. *Image Vision Comput*, 24(1), pp. 56-60.

Chapter 5

5. Sparse Reconstruction- Part II

This chapter is complementing the method of sparse reconstruction. In Chapter 4, the article concentrated only on the process of robust sparse matching and epipolar-geometry estimation. Section 5.1 explains the application of the methodology described in Chapter 4 to the gravel-pit dataset acquired with the UAV-PS. Section 5.2 describes the method of initial EO parameter estimation and sparse reconstruction. Then, Section 5.3 presents the methodology proposed for efficient sparse bundle adjustment. This Section is related to the specific objective 3.2, namely assessing a BBA strategy for on-the-job self-calibration based on the integration of pseudo-observations to Gauss-Helmert model.

5.1 Application of the proposed sparse matching technique to the gravel-pit dataset

For 159 images of dataset A (Table 3.5.1), the approximate ground fields of view (GFoV) were calculated. Then, for 639 stereo pairs, whose GFoV had more than 50% overlap, SIFT key points were detected and matched using VlFeat feature-based matching library. Then, the inlier detection and epipolar-geometry estimation were performed using the proposed algorithm, RANSAC, MSAC and MLESAC algorithms. The results are summarized in Table 5.1.1. The EO parameters from Pix4D software were used as reference data to calculate the ground-truth fundamental matrices, and to evaluate the results in terms of D_F. Where available, the checkpoints were also used to determine the accuracy of the estimated model, $\mu d.CP$. As it can be noticed, the proposed algorithm outperformed other algorithms either in terms of the computational efficiency (*Itr*) or the accuracy of model estimation (D_F, $\mu d.CP$ and μd). It should be mentioned that in these experiments, the GA iterations were set to end if the elite solutions during the stall generations were improved less than 0.1. This simple solution could handle the higher number of iterations of GA in cases the inlier ratio was high. Averagely, the proposed algorithm improved the speed of traditional RANSAC algorithm by 65%. In addition, the accuracy of fundamental-matrix estimation is improved by 61%.

Table 5.1.1. Comparative results of sparse matching for gravel-pit dataset A

Parameters	Our Alg.	RANSAC	MSAC	MLESAC
Average Itr	96	281	932	1794
Average N_{model}	2108	281	932	1794
Average VPM	1802	1835	1835	1835
mode ρ	61.1	56.7	49.4	45.1
Average μ_d	0.102	0.125	0.097	0.002
Average $\mu_d.CP$	0.399	1.020	1.914	10.439
Average D_F	12.1	50.5	24.3	136.6

5.2 Initial EO Parameter Estimation and Sparse Reconstruction

The process of SfM computation includes the following steps.

1. Establishing sparse corresponding points between stereo images
2. Estimating the fundamental matrix
3. Determining relative orientation (RO) parameters
4. Establishing a unique local coordinate system and transforming the RO parameters to EO parameters
5. Estimating the initial 3D coordinates of corresponding points
6. Transforming the local coordinate system to a global system either directly using navigation data or indirectly using the coordinates of ground control points (optional)

The presented article discussed the first step (establishing inlier corresponding points) and the second step (robust estimation of epipolar geometry). The other steps need to be accomplished as well in order to complete the process of sparse reconstruction. As a result, the initial values of EO parameters and the object coordinates of the corresponding points are generated. These estimations are used as initial values in the process of block bundle adjustment to get refined. Please note that no specific contribution was introduced in this thesis for these steps (steps 3 to 6). Therefore, the theoretical details of these steps are provided in Appendix II.

5.3 Block Bundle Adjustment

The significant problem of SfM is its relative sense of computation. SfM techniques are very useful for processing a short image sequence; however, they are not robust against the accumulation of errors in long image sequences (Kume et al., 2015). To reduce and adjust these errors, the results of SfM are optimized as initial values in a bundle adjustment (BA) process.

In BA, the estimations of exterior orientation parameters, intrinsic camera calibration parameters and 3D coordinates of the points are considered as unknowns and are optimized together. This optimization problem is usually defined as a non-linear least squares problem, where the error to be minimized is the re-projection error, i.e. the distance between an observed feature and the re-projection of the corresponding 3D point on the image. Levenberg–Marquardt (LM) has been one of the most popular least-squares optimization methods for this type of error (cost) function (Engels et al., 2006). The main approach of LM for solving a non-linear problem is to solve a sequence of linear approximations to the problem (Agarwal et al., 2010). Mostly, Gauss-Newton approach is used to build these linear approximations, called normal equations (Triggs et al., 1999).

Although BA seems computationally problematic for large-scale reconstruction, the block-sparsity of its normal equations allows its memory-efficient implementation. There is a popular trick known as Schur complementation or camera reduction, which uses this sparsity to simplify BA computationally (Triggs et al., 1999). That is the system is subdivided into structure parameters (3D coordinates of correspondences) and camera parameters. The structure ones are subsumed and a reduced system is derived that only depends on camera parameters. By solving the linear reduced system for camera parameters and back-substitution, the structure variables can be solved too. In order to solve the reduced

camera system, several solutions have been proposed in the literature. If the connectivity graph of the dataset is sparse[1] or the number of cameras is limited to a few hundred, then the most practical method for exactly and directly solving the reduced camera system is Cholesky factorization or LU decomposition[2]. SBA is one of the sparse implementations of this type of generic bundle adjustment (Lourakis and Argyros, 2009). However, if the connectivity graph is large and dense (e.g. large internet photo collections of a scene), then Cholesky factorization becomes impractical, and iterative inexact solutions based on preconditioned conjugate gradients (PCG) should be used (Byrod and Astrom, 2009; Agarwal et al., 2009). The parallel implementation of this type of BA has also been performed (Wu et al., 2011). In cases of very large datasets, the practical solution is to reduce the size of the BA problem by dividing it into sub-problems, for example by locally concentrating on the most recently modified part of the reconstruction (Mouragnon et al., 2009), by partitioning the scene to several sub-maps (Ni et al., 2007), and by partitioning the whole parameter vector into individual sub-vectors (Pang et al., 2012).

The block-bundle-adjustment strategy used in this thesis has the following characteristics and aims at achieving the objective 3.2 of this thesis.

- Non-linear Gauss-Helmert model (GHM) with Lagrange-multipliers method is used for estimation, which considers the errors in both observations and unknowns.
- Re-projection error is used to define the cost function.
- Camera system reduction is performed for sparse implementation.
- Linear reduced camera system is solved by exact, direct LU decomposition since the number of cameras in our datasets is a few hundreds, and they are not densely connected.
- Intrinsic camera calibration parameters are considered as pseudo-observations (unknowns with known weights) as explained in Section 3.6.2.

5.3.1 Sparse BBA strategy using GHM and pseudo-observations

Generally, in the literature, BA is formulated via Gauss-Markov model (GMM). Such a model is defined as $\boldsymbol{m} = \boldsymbol{f}(\boldsymbol{p})$. That is observations $\boldsymbol{m}$ (coordinates of image features) are expressed by a non-linear functional relation $\boldsymbol{f}$ of unknown parameters $\boldsymbol{p}$. The parameters include 3D coordinates of points, EO parameters and intrinsic calibration parameters of cameras. An initial estimation of parameters $\boldsymbol{p_0}$ and observations $\boldsymbol{m}$ are available from SfM computation. The final estimation of parameters $\hat{\boldsymbol{p}}$ is desired, so that the sum of squared re-projection errors $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}$ is minimized with $\boldsymbol{v} = \boldsymbol{m} - \boldsymbol{f}(\hat{\boldsymbol{p}})$. This can be solved with iterative methods such as LM. The basic key of the solver is a linear approximation of $\boldsymbol{f}$ at a small neighborhood $\delta\boldsymbol{p}$ of $\boldsymbol{p}$; i.e. $\boldsymbol{f}(\boldsymbol{p}+\delta\boldsymbol{p}) \approx \boldsymbol{f}(\boldsymbol{p})+\boldsymbol{J}\delta\boldsymbol{p}$, where $J=\partial\boldsymbol{f}/\partial\boldsymbol{p}$ is the Jacobian matrix. The main characteristic of this model is that observations would remain unchanged through the optimization iterations.

This problem can alternatively be modeled via a Gauss-Helmert model, in which the functional relation $\boldsymbol{f}$ includes both observations and unknown parameters, i.e. $\boldsymbol{f}(\boldsymbol{m}, \boldsymbol{p}) = 0$. Therefore, linearization should

[1] That is each camera looks at the same part of scene as only few other cameras

[2] LU decomposition is less efficient than Cholesky factorization for solving large linear system

be performed not only at small neighborhood $\delta\boldsymbol{p}$ of $\boldsymbol{p}$ but also at small neighborhood $\boldsymbol{v}$ of $\boldsymbol{m}$. This is the model used in this thesis and will be discussed in the following paragraphs.

The unknowns can be divided into three partitions, $\boldsymbol{p}=[\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}]\in\Re^{(6u_a+3u_b+u_c)}$. The unknowns $\boldsymbol{a_j}\in\Re^6$ represent the EO parameters of the j^{th} image. $\boldsymbol{b_i}\in\Re^3$ represents the 3D coordinates of the i^{th} point, and $\boldsymbol{c}\in\Re^{u_c}$ represents the intrinsic calibration parameters. u_a and u_b are the total number of images and object points, respectively. u_c is the number of intrinsic camera calibration parameters, which depends both on the model of distortions and the number of different cameras (devices) in the dataset. In this thesis, only one camera was used, and the IO and distortions were modeled using a total of 10 parameters (see Appendix I); i.e. u_c was equal to 10.

There may also be a set of additional constraints on unknowns as $\boldsymbol{g}(\boldsymbol{p})=0$. For instance, seven inner-constrains might be applied. As mentioned earlier, the intrinsic calibration parameters are considered as pseudo-observations; i.e. they have a weight matrix Q_c associated with them. The weight of other unknowns is zero; i.e. $Q_{p\notin c}=0$. The observations include blocks of $\boldsymbol{m_{ij}}\in\Re^2$, which represents the coordinates of the i^{th} point on the j^{th} image. The weight matrix of observations Q is made of diagonal matrices $Q_{ij}\in\Re^{2\times2}$.

With these definitions, the observation- and constraint- equations can be linearized as follows:

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{p}+\delta\boldsymbol{p},\boldsymbol{m}+\boldsymbol{v})&\approx\boldsymbol{w}+J\delta\boldsymbol{p}+B\boldsymbol{v}\\ \boldsymbol{g}(\boldsymbol{p}+\delta\boldsymbol{p})&\approx\boldsymbol{t}+M\delta\boldsymbol{p}\end{aligned}\tag{1}$$

where Jacobian matrices are $J=\partial\boldsymbol{f}/\partial\boldsymbol{p}$, $B=\partial\boldsymbol{f}/\partial\boldsymbol{m}$ and $M=\partial\boldsymbol{g}/\partial\boldsymbol{p}$, and miss-closure vectors are $\boldsymbol{w}=\boldsymbol{f}(\boldsymbol{p},\boldsymbol{m})$ and $\boldsymbol{t}=\boldsymbol{g}(\boldsymbol{p})$. Therefore, the objective function Ω to be minimized can be expressed as Equation (2), where $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are Lagrange multipliers.

$$\begin{aligned}\Omega=\boldsymbol{v}^TQ_m\boldsymbol{v}+\delta\boldsymbol{p}^{\boldsymbol{T}}Q_p\delta\boldsymbol{p}\\ +2\boldsymbol{\lambda}^T(J\delta\boldsymbol{p}+B\boldsymbol{v}+\boldsymbol{w})+2\boldsymbol{\mu}^T(M\delta\boldsymbol{p}+\boldsymbol{t})\end{aligned}\tag{2}$$

Objective function of Equation (2) would result in the following normal equations (see the proof at Appendix III):

$$\begin{bmatrix}H^c & H^{ca} & H^{cb} & 0\\ H^{ac} & H^a & H^{ab} & \left(M^a\right)^T\\ H^{bc} & H^{ba} & H^b & \left(M^b\right)^T\\ 0 & M^a & M^b & 0\end{bmatrix}\begin{bmatrix}\delta\boldsymbol{c}\\ \delta\boldsymbol{a}\\ \delta\boldsymbol{b}\\ \boldsymbol{\lambda}\end{bmatrix}=\begin{bmatrix}\boldsymbol{n}^c\\ \boldsymbol{n}^a\\ \boldsymbol{n}^b\\ -t\end{bmatrix}\tag{3}$$

where, (H^c, H^a, H^b, H^{ca}, H^{cb}, H^{ab}) are sub-blocks of Hessian matrix $H = Q_p + J^T(BQ_m^{-1}B^T)J$, ($\boldsymbol{n}^c$, $\boldsymbol{n}^a$, $\boldsymbol{n}^b$) are sub-blocks of the gradient $\boldsymbol{n} = -J^T(BQ_m^{-1}B^T)\boldsymbol{w}$, and ($M^a$, M^b) are sub-blocks of the Jacobian of constraint equations M.

Note that H^{ac}, H^{bc} and H^{ba} would be equal to $(H^{ca})^T$, $(H^{cb})^T$, $(H^{ab})^T$ only if the horizontal image coordinates were independent of vertical image coordinates. However, in this study, this condition did not hold true because of in-plane affine distortions of the sensor.

The direct solution of Equation (3) is computationally infeasible for large datasets. However, due to the sparse structure of the sub-blocks of the normal equation, it can be solved by Schur-complement trick. The following steps should be performed to solve Equation (3) sparsely. See the details of camera system reduction in Appendix III.

1. Prepare the initial estimations, $\boldsymbol{p_0}$, and image observations, $\boldsymbol{m}$ (see Section 5.2).
2. Set the termination criteria:
    a. The largest value of the residual vector $\boldsymbol{n}$ falls below ε_1 (set to 10^{-16})
    b. The largest value in the changes of parameters $\delta \boldsymbol{p}$ falls below ε_2 (set to 10^{-16})
    c. The number of LM iterations reaches an upper bound k_{max} (set to 100)
3. For any point i on image j, compute derivative matrices with respect to unknowns (J), observations (B) and constraints (M) via Equation (4).

$$
\begin{aligned}
&J_{ij}^c = \frac{\partial f_{ij}}{\partial \boldsymbol{c}},\ J_{ij}^a = \frac{\partial f_{ij}}{\partial \boldsymbol{a}_j},\ J_{ij}^b = \frac{\partial f_{ij}}{\partial \boldsymbol{b}_i} \\
&B_{ij} = \frac{\partial f_{ij}}{\partial \boldsymbol{m}_{ij}} \\
&M_i^b = \frac{\partial g}{\partial \boldsymbol{b}_i},\ M_i^a = \frac{\partial g}{\partial \boldsymbol{a}_j}
\end{aligned}
\tag{4}
$$

4. Calculate the auxiliary matrix $D_{ij} = \left(B_{ij}Q_{ij}^{-1}B_{ij}^T\right)^{-1}$
5. Compute miss-closure vectors with respect to observation equations ($\boldsymbol{w}$) and constraint equations ($\boldsymbol{t}$) via Equation (5).

$$
\begin{aligned}
&\boldsymbol{t} = \boldsymbol{g}(\boldsymbol{p}) \\
&\boldsymbol{w}_{ij} = \boldsymbol{f}_{ij}(\boldsymbol{m}_{ij}, \boldsymbol{p})
\end{aligned}
\tag{5}
$$

6. Calculate residuals with respect to camera intrinsic parameters ($\boldsymbol{n}^c$), camera orientation parameters ($\boldsymbol{n}^a$) and features 3D coordinates ($\boldsymbol{n}^b$) via Equation (6).

$$\begin{aligned}\boldsymbol{n}^c &= -\sum_{i,j}\left(A_{ij}^c\right)^T D_{ij}\boldsymbol{w}_{ij}\\ \boldsymbol{n}_i^b &= -\sum_{j}\left(A_{ij}^b\right)^T D_{ij}\boldsymbol{w}_{ij}\\ \boldsymbol{n}_j^a &= -\sum_{i}\left(A_{ij}^a\right)^T D_{ij}\boldsymbol{w}_{ij}\end{aligned} \tag{6}$$

7. Calculate Hessian sub-blocks via Equation (7).

$$\begin{aligned}&H^c = \sum_{i,j}\left(J_{ij}^c\right)^T D_{ij}J_{ij}^c + Q_c,\ H_j^{ca} = \sum_{i}\left(J_{ij}^c\right)^T D_{ij}J_{ij}^a,\ H_i^{cb} = \sum_{j}\left(J_{ij}^c\right)^T D_{ij}J_{ij}^b\\ &H_j^a = \sum_{i}\left(J_{ij}^a\right)^T D_{ij}J_{ij}^a,\ H_i^{ab} = \left(J_{ij}^a\right)^T D_{ij}J_{ij}^b,\ H_j^{ac} = \sum_{i}\left(J_{ij}^a\right)^T D_{ij}J_{ij}^c\\ &H_i^b = \sum_{j}\left(J_{ij}^b\right)^T D_{ij}J_{ij}^b,\ H_i^{ba} = \left(J_{ij}^b\right)^T D_{ij}J_{ij}^a,\ H_i^{bc} = \sum_{j}\left(J_{ij}^b\right)^T D_{ij}J_{ij}^c,\end{aligned} \tag{7}$$

8. Replace H_j^a and H_i^b with their inverses (computed using sparse LU decomposition).

9. Compute the following auxiliary variables and delete the variables that are no more required to free the memory space allocated to them.

$$\begin{aligned}&Z_i^b = H_i^b H_i^{bc},\ K^c = (H^c - \sum_i H_i^{cb}Z_i^b)^{-1},\ Z_i^g = Z_i^b K^C,\ \text{delete}\ (Z^b, H^c)\\ &K_j^O = (\sum_i H_i^{cb}H_i^b H_{ij}^{ba}) - H_j^{ca},\ \text{delete}\ (H^{ca})\\ &Z_i^d = H_i^b\left(M_i^b\right)^T,\ K^K = \sum_i H_i^{cb}Z_i^d,\ U_i^k = -Z_i^g K^K - Z_i^d,\ \text{delete}\ (Z^d)\\ &Z_i^e = H_i^b\boldsymbol{n}_i^b,\ K^N = \boldsymbol{n}^c - \sum_i H_i^{cb}Z_i^e,\ U_i^N = Z_i^e - Z_i^g K^N\\ &Z^m = -(\sum_i M_i^b U_i^K)^{-1},\ S^N = Z^m((\sum_i M_i^b U_i^N) + \boldsymbol{t}),\\ &S_j^O = Z^m((\sum_i M_i^b(-Z_i^g K_j^O - (H_i^b H_{ij}^{ba}))) + M_j^a)\end{aligned} \tag{8}$$

$$\begin{aligned}&Z_j^p = H_j^{ac}K^C,\ Z_{j_1 j_2}^o = Z_{j_1}^p\left(K_{j_2}^O + K^K S_{j_2}^O\right),\ Z^q = \left(K^N + K^K S^N\right),\ Z_j^r = Z_j^p Z^q\\ &\text{delete}\ (Z^m, Z^q, Z^p, H^{ac}, Z^e, H^{bc}, H^{cb}, H^b, \boldsymbol{n}^b, \boldsymbol{n}^c, M^b)\\ &Z_{j_1 j_2}^s = \sum_i H_{ij}^{ab}((-Z_i^g K_{j2}^O - (H_i^b H_{ij2}^{ba})) + U_i^K S_{j_2}^O),\ Z_j^u = \sum_i H_{ij}^{ab}(U_i^N + U_i^K S^N)\end{aligned}$$

$$O^c = I_{u_a} + H^a (Z^o + Z^s + (M^a)^T S^O),\ O^b = H^a (\boldsymbol{n}^a - Z^r - Z^u - (M^a)^T S^N)$$

delete $(Z^o, Z^s, Z^r, Z^u, H^a, \boldsymbol{n}^a, M^a, H^{ab})$

10. Solve the linear reduced camera system of Equation (9) using LU decomposition and update the EO parameters as $\boldsymbol{a}^{new} = \delta\boldsymbol{a} + \boldsymbol{a}$

$$O^c \delta\boldsymbol{a} = O^b \tag{9}$$

11. Calculate the Lagrange multiplier λ as in Equation (10).

$$\boldsymbol{\lambda} = S^N + S^O \delta\boldsymbol{a} \tag{10}$$

12. Solve $\delta\boldsymbol{c}$ as in Equation (11) and update the intrinsic calibration parameters as $\boldsymbol{c}^{new} = \delta\boldsymbol{c} + \boldsymbol{c}$.

$$\delta\boldsymbol{c} = K^C \left(K^O \delta\boldsymbol{a} + K^K \boldsymbol{\lambda} + K^N \right) \tag{11}$$

13. For any point *i*, solve $\delta\boldsymbol{b}_i$ via Equation (12) and update the 3D coordinates of the points as $\boldsymbol{b}_i^{new} = \delta\boldsymbol{b}_i + \boldsymbol{b}_i$.

$$\delta\boldsymbol{b}_i = U_i^N + \sum_j (-Z_i^g K_j^O - (H_i^b H_{ij}^{ba}))\delta\boldsymbol{a}_j + U_i^K \boldsymbol{\lambda} \tag{12}$$

15. Calculate the vectors of residuals $\boldsymbol{v}_{ij}$ for feature *i* on image *j* via Equation (13) and update the observation as $\boldsymbol{m}_{ij}^{new} = \boldsymbol{m}_{ij} + \boldsymbol{v}_{ij}$.

$$\boldsymbol{v}_{ij} = -(Q_{ij} B_{ij}^T)(D_{ij} (\boldsymbol{w}_{ij} + J_{ij}^c \delta\boldsymbol{c} + J_{ij}^a \delta\boldsymbol{a}_j + J_{ij}^b \delta\boldsymbol{b}_i)) \tag{13}$$

16. If any of the stop criteria (Step 2) is met, then quit. Otherwise, replace $\boldsymbol{m}$ with $\boldsymbol{m}^{new}$, replace $\boldsymbol{p_0}$ with $\boldsymbol{p}^{new} = [\boldsymbol{a}^{new}, \boldsymbol{b}^{new}, \boldsymbol{c}^{new}]$, increment *k*, and return to Step 3.

5.3.2 Experimental results

Using the observations and estimations made at Section 05.2, the BBA could be performed on images of dataset A (Table 3.5.1). The first factor tested was the effect of integrating pseudo-observations on de-correlating the unknowns. The experiments and results with this regard are already presented in Section 3.8.6, and will not be repeated here. However, as a short reminder, it was shown that the proposed strategy could de-correlate IO parameters from EO parameters up to 60%, even when the imaging configuration was not suitable.

The importance of this de-correlation becomes evident and critical when the intrinsic camera calibration parameters, estimated from BA, should be used in a separate reconstruction (intersection/resection only). This is a common case in large-scale sequential SfM, where BA can only be performed on some keyframes. In such cases, the estimated intrinsic parameters from the last BA should be directly applied for reconstructing the scene from the frames in between the key frames. Therefore, the accuracy of the intrinsic parameters would be important.

To clarify this, consider the examples given in Section 3.7.3. Once, the observations of a set of images at different altitudes (corresponding to Figure 3.7.4c) were applied for ordinary BBA and calculating the camera intrinsic parameters (corresponding to Figure 3.8.10c). Given these intrinsic calibration parameters, and some known EO parameters, the 3D coordinates of checkpoint visible on other images of the dataset were reconstructed. Note that the EO parameters were assumed fixed, and were the ones calculated from a complete BBA as in Section 3.6.2. The average 2-norm of the residuals on checkpoints in this reconstruction was 8.5±1.0 cm (blue plot at Figure 5.3.1).

Another time, the observations of a set of images at the same altitudes (corresponding to Figure 3.7.4b) were considered for ordinary BBA and calculating the camera intrinsic parameters (corresponding to Figure 3.8.10b). The average 2-norm of the residuals on checkpoints in reconstruction with these intrinsic calibration parameters was 18.3±1.8 cm (red plot at Figure 5.3.1). It should also be noted that the largest errors happened at the z-direction with an average of 18.2±1.7 cm.

One more time, the observations of the set of images at the same altitudes (corresponding to Figure 3.7.4b) were applied for the proposed BBA strategy and calculating the camera intrinsic parameters (corresponding to Figure 3.8.10d). The average 2-norm of the residuals on checkpoints in reconstruction with these intrinsic calibration parameters was only 3.5± 1.6 cm (green plot at Figure 5.3.1).

Finally, the intrinsic calibration parameters calculated offline were directly used for calculating the 3D coordinates of the checkpoints. The average 2-norm of the residuals on checkpoints in reconstruction with offline intrinsic calibration parameters was 37.1± 17.3 cm (black plot at Figure 5.3.1).

When the correlation between IO and EO parameters (specifically focal length and altitude) is high, then the IO parameters cannot be correctly estimated. Therefore, when they are applied for independent reconstruction, low accuracy is achieved. Varying the camera orientations and attitudes is the physical solution to reduce the correlation between unknowns. Therefore, the accuracy of reconstruction increases as well (53% improvement). The proposed strategy for BBA is a numerical solution to reduce the correlation between IO and EO parameters. As the results show, the accuracy of reconstruction is improved by 81%. This shows the importance of the proposed strategy in de-correlating the IO and EO parameters in bundle adjustment.

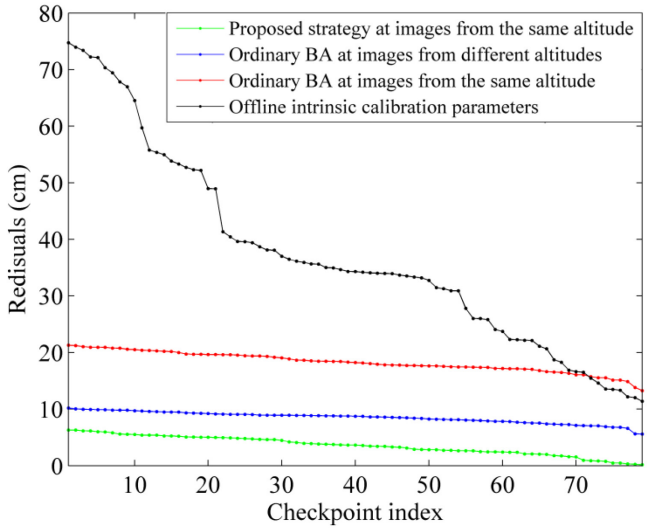


Figure 5.3.1. Error of independent reconstruction using some fixed EO parameters when applying the intrinsic calibration parameters calculated in a different way

The second effect to be tested was the behavior of the proposed sparse BBA in application to all the images of data set A. After terminating BBA, the computed coordinates of checkpoints were compared to their measured coordinates. The 2-norm of these residuals on checkpoints is presented in Figure 5.3.2. As it can be noticed, the suggested BBA strategy resulted in a high accuracy of 6.0±1.2 mm on checkpoints. Comparatively, the accuracy achieved with sparse BBA using Gauss-Markov model (see the formulation in Appendix IV) was 6.8±1.4 mm on checkpoints. It shows that, in average, the proposed strategy caused 12% accuracy improvement. In addition, the speed of convergence with the proposed strategy was almost three times faster than ordinary BA; the former converged after 5 iterations, while the latter took 15 iterations.

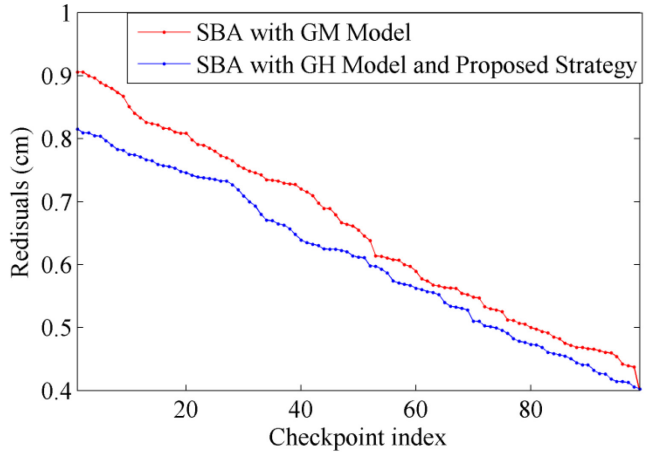


Figure 5.3.2. Accuracy of block bundle adjustment using the proposed strategy and ordinary BA

Chapter 6

6. Dense Reconstruction

6.1 Article Presentation

6.1.1 Background

This article is part of the thesis related to the specific objective 3.3, namely revisiting the concepts of intrinsic curves for developing an efficient dense stereo matching technique.

In terms of research problematic, this article considers the problem of dense matching for high-resolution images. When dealing with high-resolution images where terrain-relief variations cause large changes of disparity values, the task of dense matching becomes difficult. For instance, in the gravel-pit dataset, disparity range varied at few thousands of pixels. Therefore, the disparity search space becomes very large for any stereo pair and handling it becomes difficult either memory-wise or time-wise. Another problem with high-resolution images is that possibility of occlusion increases since more details are imaged. For instance, stereo low-resolution images of a pile might result in major occlusions due to a hidden side of the pile in one image. However, high-resolution images of the same pile result in occlusions at details too. For instance, one side of every small rock on the pile might be visible in one image but hidden in another one. In addition, the increase of details in the image increases the chance of mismatching and matching ambiguity as well. Therefore, rather than pixel-wise intensity differences, higher levels of details and clues need be considered for dense matching. Using concepts of intrinsic curves, the following article strives to address these challenges.

6.1.2 Methodology

To initiate dense matching, the stereo images should first be rectified. Rectification is the process of making vertical equivalent images from tilted stereo images (Wolf and Dewitt, 2000, p. 217). In rectified images, corresponding points are located on the same rows, namely epipolar lines or scanlines. Therefore, there is only a horizontal shift between them, called disparity. The task of dense matching is to identify the corresponding points by determining these disparity values for as many pixels as possible. In this study, method of Fusiello et al. (2000) is used for rectification, as it is known to avoid introducing large projective distortions to epipolar images. The flowchart of the research methodology for dense matching using intrinsic curves is presented in Figure 6.1.1. The following article discusses this methodology and the logic behind it. Note that this thesis is not involved with the techniques of dense matching post-refinements, such as hole-filling, densifying, sub-pixel refinement, speckle filtering of the generated disparity map, or median filtering of the generated point cloud. The objective of this article is introducing a new dense matching technique and evaluating its raw effect on the accuracy and efficiency of matching.

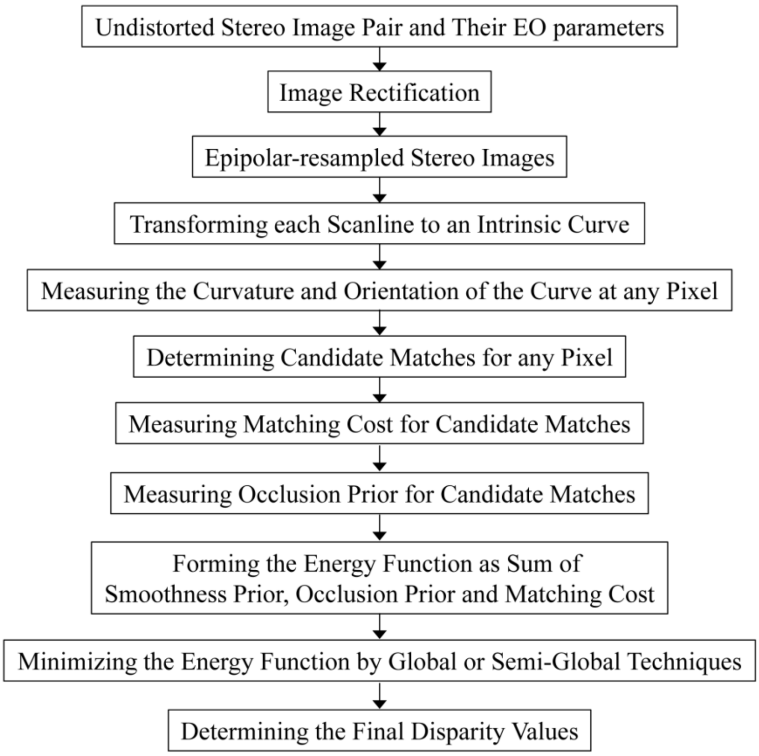


Figure 6.1.1. Methodological flowchart for dense matching using intrinsic curves

A Dense Stereo Matching Method Using Intrinsic Curves

by Mozhdeh Shahbazi, Gunho Sohn, Jérôme Théau and Patrick Ménard

ISPRS Journal of Photogrammetry and Remote Sensing,
ISSN: 0924-2716

New Submission

Abstract

Dense stereo matching is one of the fundamental and active areas of photogrammetry. The increasing image resolution of digital cameras as well as the growing interest in unconventional imaging, e.g. unmanned aerial imagery, has exposed stereo image pairs to serious occlusion, noise and matching ambiguity. This has also resulted in an increase in the range of disparity values that should be considered for matching. Therefore, conventional methods of dense matching need to be revised to achieve higher levels of efficiency and accuracy. In this paper, we present an algorithm that uses the concepts of intrinsic curves to propose sparse disparity hypotheses for each pixel. Then, the hypotheses are propagated to adjoining pixels by label-set enlargement based on the proximity in the space of intrinsic curves. The same concepts are applied to model occlusions explicitly via a regularization term in the energy function. Finally, an optimization is required to assign one of the disparity hypotheses to each pixel. With this regard, two approaches are tested, global optimization using belief propagation and semi-global matching by cost augmentation along several local paths. By searching only through a small fraction of the whole disparity search space and handling occlusions and ambiguities, the proposed framework could achieve high levels of accuracy and efficiency compared to the state-of-the-art.

Resumé

L’appariement dense est l'un des domaines fondamentaux et actifs de la photogrammétrie. La résolution croissante des images fournies par les appareils photo numériques, ainsi que l'intérêt croissant pour l'imagerie non conventionnelle telle que l'imagerie aérienne acquise par drones, occasionne des occlusions, du bruit et de l'ambigüité sur les paires d'images stéréo. Cela a également entrainé une augmentation de la gamme des valeurs de disparité qui devraient être considérées pour l'appariement. Par conséquent, les méthodes classiques d'appariement dense doivent être révisées pour atteindre des niveaux plus élevés d'efficacité et de précision. Dans cet article, nous présentons un algorithme qui utilise les concepts de courbes intrinsèques afin de proposer des hypothèses de disparité clairsemées pour chaque pixel. Ensuite, les hypothèses sont propagées aux pixels adjacents en fonction de la proximité dans l'espace des courbes intrinsèques. Les mêmes concepts sont appliqués pour modéliser les occlusions explicitement par un terme de régularisation dans la fonction d'énergie. Enfin, l'optimisation est nécessaire pour attribuer l'une des hypothèses de disparité pour chaque pixel. À cet égard, deux approches sont testées, l'optimisation globale en utilisant la propagation des croyances et l'appariement semi-globalpar l’augmentation des coûts selon divers trajets locaux. En cherchant seulement à travers une petite fraction de l'ensemble de l'espace de recherche de la disparité et la manipulation des occlusions et des ambigüités, l’approche proposée pourrait atteindre des niveaux élevés de précision et d'efficacité par rapport à l'état de l’art.

Key Words: Matching, Intrinsic Curves, Search Space Reduction, Occlusion, Energy Function, Optimization

6.2 Introduction

Dense stereo matching has always been one of the fundamental and active areas of photogrammetry and computer vision. In addition to three-dimensional (3D) scene reconstruction, several other applications such as view synthesis, image-based rendering, and robotics benefit from the results of dense matching. Generally, any dense stereo matching technique can be described by the following three components: matching cost computation, cost aggregation and disparity computation (Scharstein and Szeliski, 2002). Disparity refinement, which is basically an enhancement of the generated depth field, can also be considered as the fourth step of the dense matching.

Most of the dense matching techniques, which are discussed in the next section, still have considerable computational complexity regarding the size of the disparity search space. The techniques proposed to deal with such complexities either evaluate the complete disparity space implicitly or require successive correspondence search over the full or limited range of disparities. In addition, the sensitivity of matching techniques to occlusion, noise and matching ambiguity at poorly textured regions is undeniable. These necessitate enforcing constraints such as ordering and uniqueness that are not always efficient/true for all types of scenes.

In this paper, we present a matching algorithm that is generally based on the concepts of intrinsic curves (Tomasi and Manduchi, 1998). This matching strategy avoids exhaustive disparity search space exploration by proposing sparse disparity hypotheses for each pixel. Then, the hypotheses are propagated to adjoining pixels by label-set enlargement based on the proximity in the space of intrinsic curves in order to avoid gaps created due to noise. The same concepts are applied to model occlusions explicitly via a regularization term in the energy function. Finally, two approaches are tested to assign one of the disparity hypotheses to each pixel, global optimization using belief propagation and semi-global matching by cost augmentation along several local paths.

The rest of the paper is organized as follows. First, a brief review of literature in dense stereo matching is presented. Then, the details of the proposed technique are presented in sections 6.4 and 6.5. The experiments performed to evaluate our method are discussed in section 6.6 and, finally, the conclusions are mentioned in section 6.7.

6.3 Related Work

The techniques of dense matching can be classified into two categories of local and global methods. Local methods construct a cubic cost volume $C(x,y,d)$, which represents the cost associated with matching a pixel (x,y) in the left (reference) image to the corresponding pixel in the other stereo image (right image) at a disparity value *d* belonging to the full disparity search range. This matching cost is usually quantified by a per-pixel dissimilarity measure. Then, the disparity map can be determined by finding the minimum cost at each pixel as $\hat{d}(x,y) = \arg\min_d C(x,y,d)$. However, such results are highly noisy because the solution is not regularized. To regularize the solution one can aggregate the costs over a support region (local window) and find the disparity with the lowest aggregated cost. Basically, the

window-based cost aggregation means filtering the (x,y) dimensions of the cost volume (Hosni et al., 2013). Thus, the aggregated cost over a window W can be achieved as $C_S(x,y,d)=\sum_{(u,v)\in W}\omega_{(u,v)}C(x,y,d)$, where ω is the weight of a pixel in the support area (Gurbuz et al., 2015). For instance, in the sum-of-squared-differences (SSD) algorithm, the support aggregation is done by summing matching costs over a squared window surrounding each point assuming a constant disparity, i.e. ω=1. In some algorithms, the aggregation is performed implicitly by computing a window-based matching cost such as normalized cross-correlation (NCC). Although these window-based aggregations yield smoother results, they ignore disparity discontinuities since the windows are not aligned with image edges. In order to preserve depth discontinuities, edge-aware weighted filters, e.g. geodesic distance, bilateral and guided filtering, can be used at the cost of noticeably higher computational complexity (Yoon and Kweon, 2006; Hosni et al., 2009; Hosni et al., 2013). The main drawback of local techniques is that they require evaluating the full disparity space image (DSI). That is the matching costs should be computed and aggregated at each pixel for all possible disparities (Sinha et al., 2014). They also depend largely on the choice of the window size (W). While a small window is preferred to avoid over-smoothing and to increase computational efficiency, a large window is required in areas of low texture to decrease matching ambiguity. These facts results in a trade-off between lower success rate using small windows and border bleeding artifacts using large windows (Geiger et al., 2011). Therefore, when the local characteristics of the pixels are similar, considerable ambiguity is involved in finding their correspondences without global reasoning.

In global methods, stereo matching is formulated as a pixel labeling problem, where the inputs are a set of pixels and a set of labels (i.e. potential disparities). From the probabilistic point of view, this can be treated as an inference problem using Bayesian approaches. That is inferring the disparity map given the likelihood based on image observations and the prior based on the assumptions about the scene structure (e.g. disparity smoothness) (Sun et al., 2013). Bayesian approaches can be divided into two categories based on pathfinding and Markov random fields (MRF). In path-finding approaches, ordering and uniqueness constraints are the priors that are used to solve scanline matching in terms of finding the shortest path to go from the beginning to the end of the corresponding scanlines (epipolar lines) in the matrix of their pair-wise matching costs. The most popular solution used in the literature to find such minimum-cost path is dynamic programming (DP) (Cox et al., 1996). The main drawback of these techniques is that the dependence between scanlines is either totally ignored or partially considered via inter-scanline vertical edges (Ohta et al., 1985). Without the smoothness assumption between scanlines, the results of path-finding approaches often suffer from streaking effect (Zitnick and Kanade, 2000). As a powerful alternative, the labeling problem can be modeled as a Markov random field (MRF) since a particular pixel label depends only on the labels of its neighbors (Boykov et al., 1998). Such problem can then be solved in an energy minimization framework, where a global energy function penalizes intensity dissimilarities between the corresponding pixels and discontinuities between the neighboring pixels in the disparity map (Scharstein and Szeliski, 2002). In other words, the maximum a posterior (MAP) estimate of the disparity map can be achieved by minimizing this energy function. However,

minimizing a global MRF-based energy function is generally NP-hard[1]. Therefore, a variety of approximation algorithms has been proposed that apply graph cuts or belief propagation for inference (Sun et al., 2003; Boykov et al., 2001). Semi-global matching (SGM) is also a technique that approximates the global MRF inference by aggregating cost functions along several (usually eight or sixteen) local paths in the image and computes the disparity with a winner-take-all mechanism (Hirschmüller, 2008).

A drawback of global methods is their computational complexity, which does not scale well to large label spaces (large DSI). Several solutions exist for limiting the disparity search range. The simplest way is to select only the best disparities for each pixel that correspond to the highest matching scores. Another way is the hierarchical approach, in which a Gaussian pyramid of the original images is constructed and disparities are computed at each level. Then, disparity results at coarser levels are used to reduce the disparity range at finer levels. Local, fast window-based techniques can also be used before applying a complex global technique to reduce the matching ambiguity. A more recent category of techniques for disparity search space reduction is based on assuming planar hypotheses for specific regions of the images. For instance, Libelas method builds a prior on the disparities by forming a triangulation on sparsely matched keypoints (Geiger et al., 2011). PatchMatch stereo also finds correspondences between small patches (segments) of the images by iteratively propagating disparities from initial seeds to their neighbors (Bleyer et al., 2011). In LPS (Local Plane Sweeps) method, local slanted plan hypotheses, which are derived from initial sparse feature correspondences, are used to propose disparity hypotheses (Sinha et al., 2014). Except for the best candidate based method, the rest of these techniques require either exhaustive DSI computation or successive matching to find the appropriate priors.

Another drawback of global methods is their tendency to fail in occluded regions (Mozerov et al., 2015). Generally, uniqueness and ordering constraints are used to handle occlusion. Uniqueness is usually enforced by left-right cross-checking of disparity maps. The ordering constraint, which is mostly satisfied in DP-based methods, requires that the relative ordering of pixels remains the same on the corresponding scanlines. However, such assumption is not always true, especially in scenes containing narrow foreground objects (Scharstein and Szeliski, 2002).

6.4 Original Concepts of Intrinsic Curves

6.4.1 Definition

In this study, we revisit the concept of intrinsic curves proposed by Tomasi and Manduchi (1998) to develop a new dense matching approach. Intrinsic curves are the multidimensional representations of the paths that the image descriptors follow as a scanline is traversed from left to right. Assume the intensity and its derivative (gradient) as two image descriptors. Scanlines can then be considered as one-dimensional (1D) signals of intensity $l(x^l)$ at every location x^l on the left scanline and $r(x^r)$ at every

[1] non-deterministic polynomial-time hard

location x^r on the right scanline (Figure 6.4.1a)[1]. Respectively, their derivatives are $l'(x^l)$ and $r'(x^r)$. If we plot $l'(x^l)$ versus $l(x^l)$, then we lose the spatial track of x^l (Figure 6.4.1b), i.e. the new representation of the left scanline would be $C^l=l'(l)$. In this new representation, if $l(x)$ is replaced by a displaced replica $l(x^l+d)$, the curve C^l of Figure 6.4.1b remains the same. Due to this invariance to displacements (disparities), these curves are called intrinsic curves. Then, the problem of image matching would be converted to the problem of curve matching. It should be noted that in order to have a continuous/smooth representation of the curves, the scanline intensity signals are modeled with piece-wise cubic splines with regard to x.

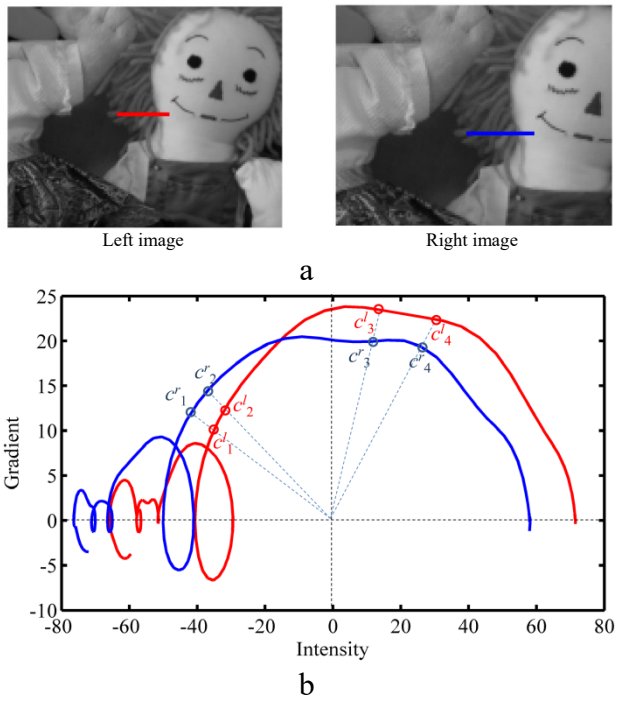


Figure 6.4.1. (a) One part of left and right corresponding scanlines; (b) Intrinsic curves of the left and right scanlines (C^l, C^r) in red and blue, respectively.

6.4.2 Matching cost computation

If the only difference between two scanlines were the geometric disparity, the two curves would coincide everywhere. However, the images are usually corrupted by noise and photometric transformations. Therefore, there is a non-constant variation between the two curves. According to Tomasi and Manduchi (1998), zero-mean low-pass filtering of the images may remove the noise and brightness bias between images completely. Therefore, contrast difference is the only remaining difference between scanlines, which causes the right intrinsic curve C^r be an expanded/contracted form of the left intrinsic curve C^l from the origin. This assumption suggests a radial metric for finding candidate matches based on intrinsic curves; i.e. two points might be corresponding if they are collinear with the origin in the two-dimensional space of the curves, e.g. points $\boldsymbol{c}^l_1$ and $\boldsymbol{c}^r_1$ in Figure 6.4.1b. In this

[1] The stereo images belong to Middlebury Stereo Datasets; see Section 5.6.

sense, their radial distance in the full space of curves shows the degree of their similarity and measures the matching costs.

6.4.3 Local aggregation

The original study suggests aggregating the candidate matches into candidate matching segments; i.e. a candidate match and its close neighbors on the intrinsic curves belong to a matching segment, and these matches are either all wrong or right. For instance, candidate correspondences ($\boldsymbol{c}^l_1$, $\boldsymbol{c}^r_1$) and ($\boldsymbol{c}^l_2$, $\boldsymbol{c}^r_2$) belong to the same segment since the arc-length s_l between $\boldsymbol{c}^l_1$ and $\boldsymbol{c}^l_2$ is smaller than a threshold and is also close to the arc-length s_r between $\boldsymbol{c}^r_1$ and $\boldsymbol{c}^r_2$. The matching cost of each segment is defined as the sum of the matching costs of its correspondences.

6.4.4 Disparity computation

Because of the aggregation step, there are fewer candidate segments than candidate matching points; therefore, the search space for matching is reduced. Constraints of uniqueness and ordering are applied to the candidate segments to solve the matching with a path-finding approach. That is, two candidate segments, e.g. $S_1=\{(\boldsymbol{c}^l_1,\boldsymbol{c}^r_1),(\boldsymbol{c}^l_2,\boldsymbol{c}^r_2)\}$ and $S_2=\{(\boldsymbol{c}^l_3,\boldsymbol{c}^r_3),(\boldsymbol{c}^l_4,\boldsymbol{c}^r_4)\}$, can follow each other only if there is no overlap between them (uniqueness) and one is the successor of the other (ordering). Based on this concept, a graph is formed in which the candidate segments are the nodes, and segments that can follow each other are linked with edges. The edges are weighted based on the aggregated matching cost of the source nodes. Then, an application of a shortest-path-finding algorithm produces the minimum-cost path through the scanline, that is, the best matching segments.

6.4.5 Shortcomings of the original concepts

Although low-pass filtering of a signal reduces its noise level and subtracting the mean from two signals decreases their shift (bias), there usually remains some local shift between the signals. In other words, assuming an affine photometric distortion model is locally possible, but not globally. As a result, the radial metric (phase similarity) is not applicable everywhere. For example, at the stereo pair of Figure 6.4.1a, only for 35% of the points, the ground-truth match exists among the hypothesized candidate matches.

In addition, computing matching cost as the radial distance between two points is not enough to resolve the matching ambiguities since this distance is only a pixel-wise measure, similar to squared intensity difference (SD).

Another shortcoming arises from the aggregation step. While segmenting candidate matches based on the arc-length proximity seems like a brilliant alternative for window-based cost aggregation, it does not provide dense matching and struggles for finding correct matches at poorly textured image areas. To understand this more clearly, a part of a scanline with both poor and high-variance textures is shown in Figure 6.4.2. Sampling the intensity signal of this scanline based on a constant x (1 pixel) results in a uniform grid through the scanline (Figure 6.4.2a). However, sampling based on a constant arc-length (9 gray values) through the intrinsic curve results in a non-uniform grid through the scanline (Figure 6.4.2b). This non-uniform grid is denser at busy areas of the image since the arc-lengths are longer where high intensity changes happen. As a result, at these areas candidate matches may not be close

enough (in terms of arc-length) to form candidate segments; therefore, they will be left unmatched. On the other hand, at flat areas of the image, the arc-lengths are shorter. Thus, these areas produce crowded candidate segments with high ambiguity that result in wrong matches.

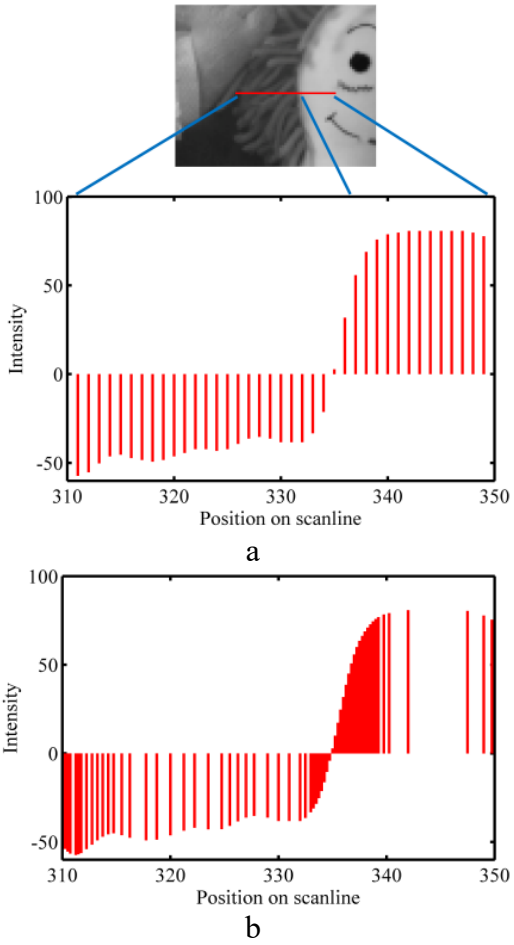


Figure 6.4.2. (a) Uniform sampling of a scanline based on position; (b) Non-uniform sampling based on arc-length.

6.5 Proposed Dense Matching Algorithm

In contrast with the original concept, we do not want to lose the full track of disparity. However, we are looking for a solution to maximize the benefits from the concepts of intrinsic curves in order to first, reduce the disparity search space without requiring successive matching or additional search, and, second, to handle the poorly-textured areas (ambiguity) and occluded areas (occlusion) efficiently.

6.5.1 Hypothesis generation

We establish the hypotheses for pixels belonging to a small neighborhood (N_i) by assuming that the photometric transformation between corresponding scanlines can be locally modeled with an affine transformation as follows:

$$r(x^l + d) = \alpha_i l(x^l) + \beta_i \quad , x^l \in N_i \tag{1}$$

where the gain parameter α_i and the shift parameter β_i represent the difference in contrast and brightness between the scanlines in the local neighborhood N_i. This assumption necessitates zero-mean low-pass filtering of the image, e.g. with a Gaussian filter with kernel size 3 and standard deviation of 1.2, in order to avoid an additional noise term in Equation (1). In fact, the noise term can be considered small enough to be independent of x and to be integrated to the brightness bias term.

If we consider the parametric representation of the left intrinsic curves by a pair of functions as $C^l(x) = (l(x), l'(x))$, then the parametric form of the right curve can be estimated as $C^r(x^l + d) = (\alpha_i l(x^l) + \beta_i, \alpha_i l'(x^l))$. This proves that, at corresponding points, C^l and C^r cannot be radial from the center; however, the tangents to the curves, $\boldsymbol{t}^l = (l', l'')$ and $\boldsymbol{t}^r = (r', r'')$, have the same orientations (θ^l, θ^r).

$$\begin{aligned} \theta^l(x^l) &= \tan^{-1}\left(\frac{l''}{l'}\right) \\ \theta^r(x^l + d) &= \tan^{-1}\left(\frac{r''}{r'}\right) = \tan^{-1}\left(\frac{\alpha_i l''}{\alpha_i l'}\right) \end{aligned} \tag{2}$$

Therefore, at any pixel $\boldsymbol{p}$ at coordinates (x_p, y_p) on the left image, the set of disparity candidates D_p can be defined as follows:

$$D_p = \{d \in D \mid abs(\theta^l(x_p) - \theta^r(x_p + d)) < T\} \tag{3}$$

where D is the whole range of possible disparities.

Although this assumption is generally true for small thresholds T, it may cause problems at texture-less areas of the image, where intensity changes are either very low or zero. Because, in these cases, even small values of remaining noise on the scanlines are large enough, compared to the values of (l', l''), to make noticeable differences between θ^l and θ^r. To avoid such situations, we can increase the threshold T. In addition, a filtering step is applied to fill in any gaps in the disparity hypotheses that may exist because of noise. This filtering step is similar to the idea proposed by Veksler (2006), however with different definition and measures. Let define set P_d as $P_d = \{\boldsymbol{p} \mid d \in D_p\}$. In other words, P_d is the set of all the pixels for which the disparity d is selected as a candidate disparity. Now, P_d can be extended (enlarged) to all the pixels that are close to pixels of set P_d on the intrinsic curves. That is:

$$P_d^{extended} = \{\boldsymbol{p_i} \mid \varphi(\boldsymbol{p_i}, \boldsymbol{p_j}) < \sigma \ \text{ for some } \boldsymbol{p_j} \in P_d\} \tag{4}$$

where $\varphi(\boldsymbol{p_i}, \boldsymbol{p_j})$ is the curve arc-length measured as:

$$\varphi(\boldsymbol{p_i}, \boldsymbol{p_j}) = \int_{xp_i}^{x_{p_j}} \left((l'(x))^2 + (l''(x))^2\right)^{1/2} dx \tag{5}$$

It is noteworthy that this measure to extend the disparity candidates considers the fact that close pixels (in spatial space) that have similar intensities and similar intensity changes (therefore, are close in the curve space) are more probable to have similar disparities.

6.5.2 Occlusions

Using intrinsic curves, occlusions stand out as pieces of one curve, usually in the form of loops, which remain unmatched in the other curve. Figure 6.5.1a and Figure 6.5.1b show two scanlines and their corresponding intrinsic curves. The stars on the curves show the matched pixels and the circles on the left curve show the occluded pixels.

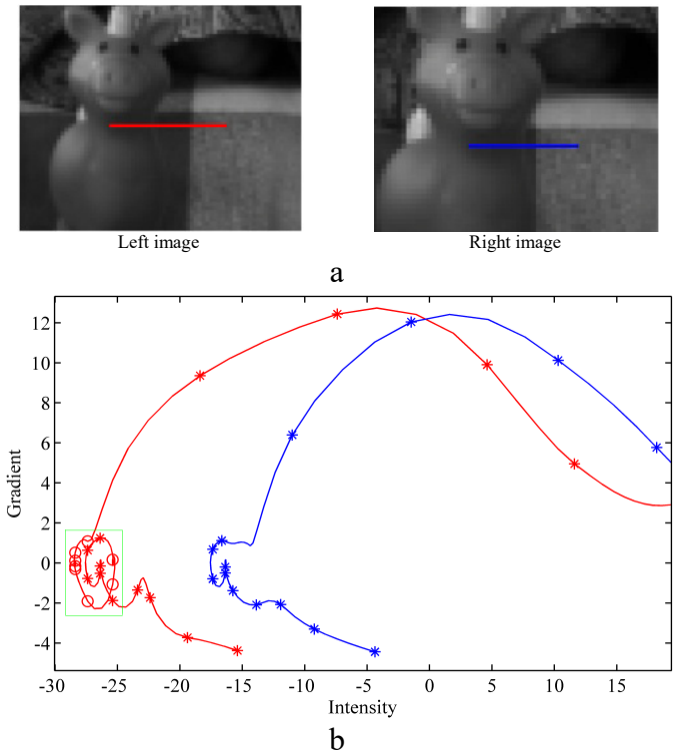


Figure 6.5.1. (a) One part of left and right corresponding scanlines with partial occlusions; (b) Intrinsic curves demonstrating matched pixels (shown by stars) and occluded pixels (shown by circles in the green frame).

As it can be noticed, occlusions happen at left arcs whose curvature changes are not similar to the right arcs. In other words, the curvature of the left curve changes the same way as the curvature of the right curve between any two non-occluded corresponding points; i.e. the curve remains either concave or convex in both curves. To formulate this concept mathematically, consider a pixels x^l_i on the left scanline and its adjacent neighbors x^l_i+1 and x^l_i-1. Their correspondences on the right scanline are denoted as x^r_i, x^r_i+1 and x^r_i-1. Respectively, the points on the intrinsic curves associated with these pixels are denoted as $\boldsymbol{c}^l_i$, $\boldsymbol{c}^l_{i+1}$, $\boldsymbol{c}^l_{i-1}$, $\boldsymbol{c}^r_i$, $\boldsymbol{c}^r_{i+1}$ and $\boldsymbol{c}^r_{i-1}$ (Figure 6.5.2). Assume that ${}^l\kappa_i^{i+1}$ contains the

curvature values of the left curve between $\boldsymbol{c}^l_i$ and $\boldsymbol{c}^l_{i+1}$. A similar definition applies to ${}^l\kappa^{i}_{i-1}$, ${}^r\kappa^{i+1}_{i}$ and ${}^r\kappa^{i}_{i-1}$. Therefore, the similarity of curvature changes around pixels x^l_i and x^r_i can be measured as follows.

$$\Theta(x^l_i, x^r_i) = \min_{j\in\{i,i-1\}} \left(\left| sign({}^l\kappa^{j+1}_{j}) - sign({}^r\kappa^{j+1}_{j}) \right| \right) \tag{6}$$

If x^l_i is non-occluded and is surrounded by two non-occluded pixels x^l_i+1 and x^l_i-1, then $sign({}^l\kappa^{i+1}_{i})$ and $sign({}^r\kappa^{i+1}_{i})$ are the same; besides, $sign({}^l\kappa^{i}_{i-1})$ and $sign({}^r\kappa^{i}_{i-1})$ are equal too, which means $\Theta(x^l_i, x^r_i) = 0$. Alternatively, if it is surrounded by one occluded pixel at one side, for example at x^l_i+1, and by one non-occluded pixel at the other side (x^l_i-1), then only $sign({}^l\kappa^{i}_{i-1})$ and $sign({}^r\kappa^{i}_{i-1})$ are equal; however, $sign({}^l\kappa^{i+1}_{i})$ and $sign({}^r\kappa^{i+1}_{i})$ are different. This shows an occurrence of occlusion. This concept is used later in defining the global energy function for matching.

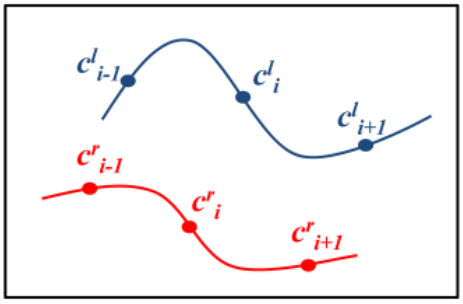


Figure 6.5.2. Example of intrinsic curves to measure local curvature similarity for occlusion detection

It should be noted that occlusions are also implicitly considered in the stage of hypothesis generation. That is, the pixels with empty candidate disparity sets are automatically recognized as occluded pixels and will be left unmatched.

6.5.3 Global energy function

Consider the reference image as a set of pixels P with observed intensities I_p for each $\boldsymbol{p} \in P$. Therefore, a pixel $\boldsymbol{p}$ is representative of image observations at pixel $\boldsymbol{p}$. The pixel $\boldsymbol{p}$ has coordinates (x_p,y_p) and its four immediate neighbors $\{(x_p-1,y_p), (x_p+1,y_p), (x_p,y_p-1), (x_p,y_p+1)\}$ form a neighborhood set N_p. The set of all candidate disparities D is defined as $D = \bigcup_{\boldsymbol{p}\in P} D_p$, where D_p is the set of disparity hypotheses for pixel p as defined in Section 6.5.1.

In global algorithms, the goal of matching is to compute, for each pixel $\boldsymbol{p}$, an optimal label d_p as the disparity at this pixel, so that the following energy function is minimized:

$$E(P,D)=\sum_{\boldsymbol{p}\in P}\left\{E_{data}(\boldsymbol{p},d_p)+E_{occ}(\boldsymbol{p},d_p)+\sum_{\boldsymbol{q}\in N_p}E_{smooth}(d_p,d_q)\right\} \tag{7}$$

In the following subsections, the data term E_{data} and smoothness term E_{smooth}, the occlusion term E_{occ} are defined. Later in Section 6.5.4, the belief propagation strategy to minimize this energy function is explained as well. In addition, in Section 6.5.5, semi-global matching using these occlusion and data terms is explained.

6.5.3.1 Data term

The data term encodes the intensity similarity (photometric consistency) of pixel correspondences for hypothesized disparities:

$$E_{data}(\boldsymbol{p},d_p)=\rho_{data}(F(\boldsymbol{p},d_p)) \tag{8}$$

where $F(\boldsymbol{p},d_p)$ is the cost of matching pixel $\boldsymbol{p}$ to the right image with disparity d_p. In this study, we measure F based on non-parametric census transform (Zabih and Woodfill, 1994). The function ρ_{data} is a truncated L_1 norm function proposed by (Sun et al., 2003) that is robust to noise:

$$\rho_{data}(t)=-\ln\left((1-e_d)\exp\left(-\frac{|t|}{\sigma_d}\right)+e_d\right) \tag{9}$$

where e_d and σ_d are parameters controlling the shape of the function.

6.5.3.2 Smoothness term

The smoothness term encodes the piecewise smoothness prior on the disparities of every pixel $\boldsymbol{p}$ and its immediate neighbors $\boldsymbol{q}$.

$$E_{smooth}(d_p,d_q)=\rho_{smooth}(d_p-d_q) \tag{10}$$

where the function ρ_{smooth} is also a robust function defined similar to Equation (9) with parameters e_s and σ_s.

$$\rho_{smooth}(t)=-\ln\left((1-e_s)\exp\left(-\frac{|t|}{\sigma_s}\right)+e_s\right) \tag{11}$$

The function ρ_{smooth} has the form of a potential function of Total Variance (TV), which has the discontinuity-preserving property needed for a proper smoothness term (Sun et al., 2003). The parameter e_s controls the upper bound (truncation) of the model and the parameter σ_s defines the sharpness (Figure 6.5.3).

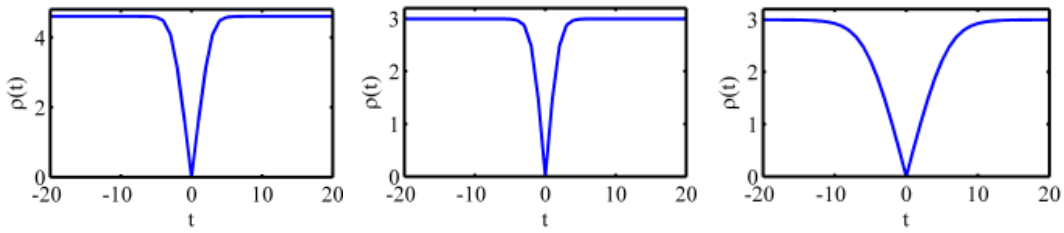


Figure 6.5.3. The robust function $\rho(t)$ with different parameters. (a) e=0.01, σ=0.6; (b) e=0.05, σ=0.6; (c) e=0.05, σ=1.8.

6.5.3.3 Occlusion term

In addition, we want to include the occlusion term for encoding the occlusion clues as explained in Section 6.5.2. Therefore, we integrate such occlusion assumption into the basic energy function using a soft constraint as E_{occ}.

$$E_{occ}(\boldsymbol{p}, d_p) = \Theta(x_p, x_p + d_p) \tag{12}$$

where the function Θ is defined as in Equation (6). This energy term imposes a penalty for a pixel $\boldsymbol{p}$ being occluded on the right image assuming a disparity d_p.

6.5.4 Approximating inference by belief propagation

From probabilistic point of view, the energy is equal, up to a constant, to the negative log posterior. Therefore, minimizing the energy function $E(P,D)$ is equivalent to maximizing a posterior probability $p(D|P)$; that is the disparity map given image observations (Equation (13)). In fact, in a Markov random field network, pixels P can be considered as observable nodes of the graph, and disparity hypotheses D (or labels) can be considered as random variables (hidden nodes). There is a link (undirected edge) between any pixel $\boldsymbol{p}$ and all the labels in its candidate disparity set D_p and also between any directly neighboring pixels:

$$posterior \propto \prod_{\boldsymbol{p} \in P} \left(\gamma(\boldsymbol{p}, d_p) \prod_{\boldsymbol{q} \in N_p} \eta(d_p, d_q) \right) \tag{13}$$

where γ is the unary potential,

$$\gamma(\boldsymbol{p}, d_p) = \exp\left(-\rho_{data}(F(\boldsymbol{p}, d_p) - \Theta(x_p, x_p + d_p)\right) \tag{14}$$

and η is the interaction or binary potential.

$$\eta(d_p, d_q) = \exp(-\rho_{smooth}(d_p - d_q)) \tag{15}$$

To maximize this posterior, loopy belief propagation (LBP) technique is used (Sun et al., 2003). There are several algorithms to perform LBP, from which the max-product algorithm is applied here that maximizes the joint posterior. Note that this would be equivalent to the min-sum algorithm, if the negative log probabilities were used instead of the potentials. Max-product is basically a message-

sending algorithm where every observable node $\boldsymbol{p}$ sends a message ($msg_{p\to q}(d)$) to node $\boldsymbol{q}$ in its neighborhood about the amount of its belief that node $\boldsymbol{q}$ has disparity value d:

$$msg_{\boldsymbol{p}\to\boldsymbol{q}}(d)=\max_{l\in D_p}\left(\gamma(\boldsymbol{p},l)\eta(d,l)\prod_{\boldsymbol{s}\in N_p\backslash\boldsymbol{q}}msg_{\boldsymbol{s}\to\boldsymbol{p}}(l)\right) \tag{16}$$

where $N_p\backslash\boldsymbol{q}$ means all the pixels in the neighborhood of pixel $\boldsymbol{p}$ except for $\boldsymbol{q}$, to which the message is being sent. It should be noted that for a pixel $\boldsymbol{q}$ should only receive messages about disparities that belong to its candidate disparity set, i.e. $d\in D_q$.

The details of the algorithm are avoided here and readers are referred to Sun et al. (2003) and (2005) for more details. The message sending iterates several times until all the nodes receive the complete messages from the other nodes. Then, at the end, the belief at each node $\boldsymbol{p}$ about any disparity candidate d (i.e. $B_p(d)$) can be computed as follows.

$$B_p(d)=\gamma(\boldsymbol{p},d)\prod_{\boldsymbol{q}\in N_p}msg_{\boldsymbol{q}\to\boldsymbol{p}}(d) \tag{17}$$

Therefore, the disparity candidate which maximizes the belief is the optimal one ($\hat{d}_p$), which together with the optimal disparities at other pixels maximizes the posterior of Equation (13), or equivalently, minimizes the energy of Equation (7).

$$\hat{d}_p=\arg\max_{d\in D_p}\left(B_p(d)\right) \tag{18}$$

6.5.5 Semi-global optimization

As explained in Section 6.3, semi-global matching is a technique that approximates the global MRF inference by aggregating pixel-wise cost functions and smoothness priors along several local paths in the image. Then, it determines the disparity with a winner-take-all mechanism. In this study, the energy function is defined as Equation (7), with a slight change for the smoothness prior. Based on the original method of SGM (Hirschmüller, 2008), the following energy function is adapted,

$$E(P,D)=\sum_{\boldsymbol{p}\in P}\left\{\rho_{data}(F(\boldsymbol{p},d_p))+\Theta(x_p,x_p+d_p)+\sum_{q\in N_p}\left(P_1T\left[\left|d_p-d_q\right|=1\right]+P_2T\left[\left|d_p-d_q\right|>1\right]\right)\right\} \tag{19}$$

where T is a Boolean operator evaluating to 1 if the succeeding condition is true and to 0 otherwise. Values P_1 and P_2 add constant penalties to the energy when disparity changes a little bit and largely, respectively. Therefore, it should always be ensured that $P_2\geq P_1$.

To semi-globally minimize the energy function of Equation (19), the aggregated costs $S(\boldsymbol{p},d)$ should be computed. To this end, pixel-wise costs should recursively be accumulated along several paths $\boldsymbol{r}$, e.g. $\boldsymbol{r}$=(-1,1), as follows,

$$L_r(\boldsymbol{p},d_p)=\rho_{data}(F(\boldsymbol{p},d_p))+\Theta(x_p,x_p+d_p)+\min\begin{Bmatrix}L_r(\boldsymbol{q},d_p)\\L_r(\boldsymbol{q},d_p-1)+P_1\\L_r(\boldsymbol{q},d_p+1)+P_1\\\min\limits_{i\in D_q\setminus\{d_p,d_p-1,d_p+1\}}L_r(\boldsymbol{q},i)+P_2\end{Bmatrix}-\min_{i\in D_q}L_r(\boldsymbol{q},i) \tag{20}$$

where $\boldsymbol{q}=\boldsymbol{p}+\boldsymbol{r}$. Notice that the terms $L_r(\boldsymbol{q},d_p)$ exist only if $d_p\in D_q$. Therefore, the path accumulation is performed only for $d_p\in D_p$ and the terms that involve $d_p\notin D_q$ are set to infinity so that they do not affect the minimizing operator in Equation (20). Afterward, the accumulated costs L_r are summed over the paths in all directions of r, e.g. 8 paths.

$$S(\boldsymbol{p},d_p)=\sum_{\boldsymbol{r}}L_r(\boldsymbol{p},d_p) \tag{21}$$

Finally, the optimal disparity value is selected as the one that minimizes the aggregated cost S.

$$\hat{d}_p=\arg\min_{d_p\in D_p}S(\boldsymbol{p},d_p) \tag{22}$$

6.6 Experimental Results

6.6.1 Close-range stereo images

In order to test different aspects of the proposed dense matching algorithm, some training stereo images belonging to the Middlebury Stereo benchmark were used (Scharstein and Szeliski, 2003; Scharstein and Pal, 2007; Hirschmüller and Scharstein, 2007; Scharstein et al., 2014). Moreover, evaluations were performed on the 2014 test datasets of the benchmark, which allowed the comparison of this algorithm with other state-of-the-art ones.

For hypothesis generation, orientation threshold of T=15 degrees was used (in Equation (3)), and the maximum arc-length in the curve was considered to decide the arc-length threshold σ in Equation (4). If the number of pixels, for which at least one disparity candidate was generated, was less than 40% of the image size, the orientation threshold was augmented to T=45. However, it was ensured that the number of candidates for other pixels did not increase. For F in Equation (8), non-parametric census transform with a window size of 9×7 pixels was used. For data and smoothness terms in Equations (9) and (11), e_s=0.15, σ_s=0.6, e_d=0.01 and σ_d=3 were selected. For LBP, the maximum number of iterations was set to 25, and for SGM, the penalty terms P_1=27 and P_2=29 were used, and the number of paths was set to 8 (equivalent to 4 paths in two scans).

In the following experiments, the variables that were evaluated include the followings.

- gain in computational efficiency in terms of reducing the size of the disparity search space
- accuracy of hypothesis generation to initiate the matching

- effect of the inclusion of occlusion term in the energy function
- accuracy of one-way matching (without any left-right cross consistency checking) compared to the state-of-the-art
- raw accuracy of matching (after left-right consistency checking, with no other post-processing) compared to the state-of-the-art
- density of the generated disparity maps after dense matching

To evaluate these variables, several performance criteria were used that are described in Table 6.6.1.

Table 6.6.1. Description of performance-assessment criteria

Symbol	Description
1-way error	Number of corresponding points whose disparity values are more than 1.5 pixels different from the ground-truth disparity values, divided by the total number of matched pixels, multiplied by 100 It is calculated from the disparity map obtained by directly matching the left image to the right one.
2-way error	Similar to the 1-way error However, it is calculated from the disparity map obtained after left-right cross-consistency checking
Density	Total number of matched pixels divided by the total number of matches in the ground-truth map, multiplied by 100
Hyp_Acc	Total number of pixels for which the ground truth disparity value is among their hypothesized candidate disparities, divided by the total number of matches in the ground-truth map, multiplied by 100
Hyp_Size1	Number of candidate disparities hypothesized for each pixel averaged over all the pixels for which at least one disparity candidate is hypothesized, divided by the ground-truth disparity range of the image, multiplied by 100
Hyp_Size2	Number of candidate disparities hypothesized for each pixel averaged over all the pixels for which at least one disparity candidate is hypothesized, divided by the image width (which is the disparity search range if no information is available beforehand), multiplied by 100

Several variants of our proposed algorithm and other state-of-the-art techniques of dense matching were applied. Unless otherwise mentioned, for the state-of-the-art techniques the open-source/access libraries provided by their authors and the same parameters mentioned in their corresponding articles were used. Table 6.6.2 presents a short description of these techniques.

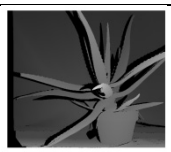
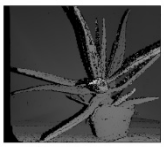
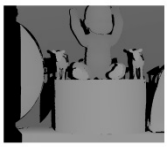
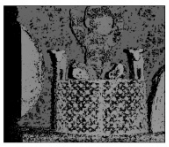
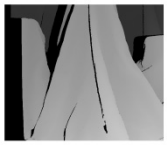
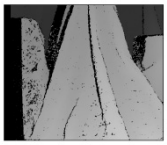
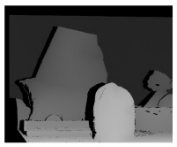
Table 6.6.2. Description of dense matching techniques applied for comparative experiments

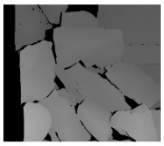
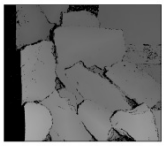
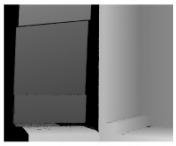
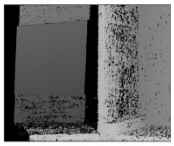
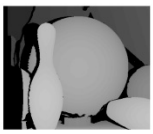
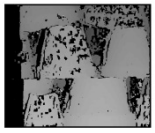
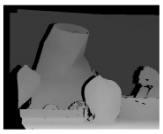
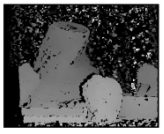
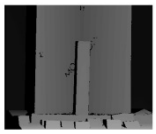
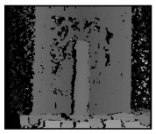
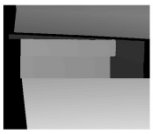
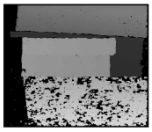
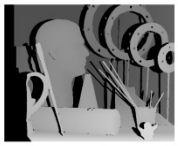
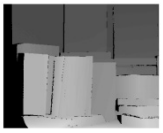
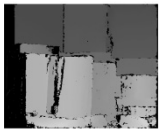
Technique Symbol	Description	Reference
ICBP	The proposed algorithm of dense matching based on intrinsic curves, with final LBP optimization as described in Section 6.5.4	This paper
ICSG	The proposed algorithm of dense matching based on intrinsic curves, with final semi-global optimization as described in Section 6.5.5	This paper
ICWTA	The proposed algorithm is used for disparity hypothesizing and measuring total matching cost as the sum of data cost, Equation (9), and occlusion term, Equation (12). Final matching is done by winner-take-all mechanism.	This paper
ICBP_Noocc	The same as ICBP, except that the occlusion term, Equation (12), is not included in the energy function.	This paper

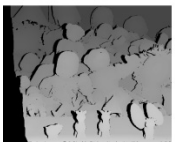
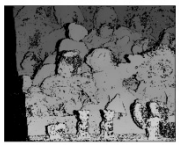
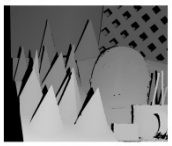
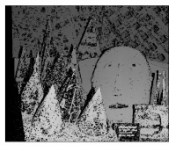
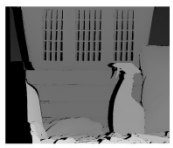
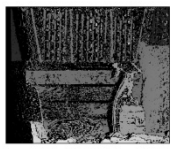
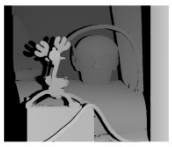
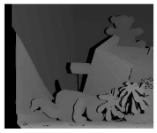
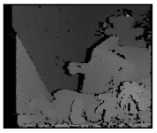
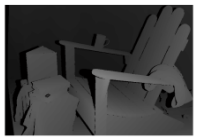
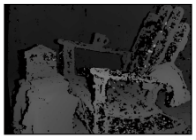
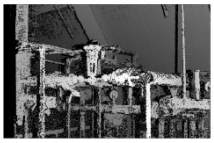
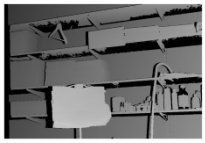
ICSG_Noocc	The same as ICSG, except that the occlusion term, Equation (12), is not included in the energy function.	This paper
ICWTA_Noocc	The same as ICWTA, except that the occlusion term, Equation (12), is not included in the matching cost.	This paper
CenWTA	Local matching using 9 × 7 Census cost.	
RSGM	A CPU implementation of conventional SGM with fine-grained parallelization and usage of multiple cores, where the disparity space is partially sub-sampled and compressed. 9×7 Census mask is used as input for the matching cost calculation. The range of disparity search space is limited to 256 pixels.	Spangenberg et al., 2014
ADGC	Correspondences with occlusions are calculated using graph cuts. The measure of matching cost is the absolute difference of intensity that is calculated by linearly interpolating intensity values surrounding the pixels.	Kolmogorov and Zabih, 2001
ADHierBP	Matching is performed by hierarchical (multi-level) loopy belief propagation, which reduces the number of message-sending iterations at lower levels (higher resolutions). The matching cost is the same as ADGC.	Felzenszwalb and Huttenlocher, 2006
SURE	Matching is performed using a hierarchical coarse-to-fine solution for the conventional SGM method, in which matching results of low-resolution images are used to narrow down the disparity search range for higher resolutions of the pyramid. As the cost of matching, 9 × 7 Census cost is used.	Rothermel et al., 2012

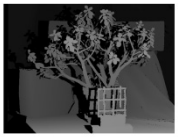
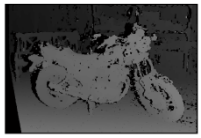
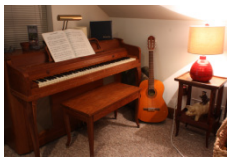
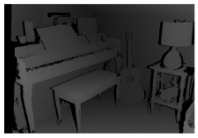
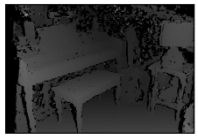
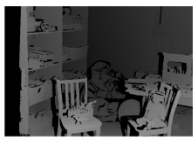
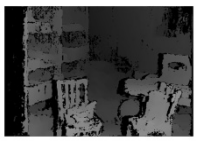
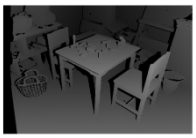
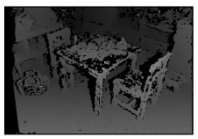
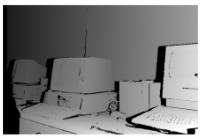
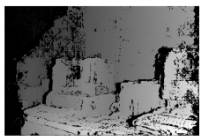
Table 6.6.3 presents the data used for the close-range stereo experiments. For each dataset, the left image, the ground-truth disparity map, the estimated disparity map by ICBP, the parameters Hyp_Size1, Hyp_Size2 and Hyp_Acc (in percentage) are illustrated. In this table, the capital letter *L* at the beginning of dataset name stands for low-resolution, and *H* stands for high-resolution.

Table 6.6.3. Test datasets, ground-truth and calculated disparity maps and performance of the proposed hypothesis-generation algorithm

Dataset Name	**Left Image**	**Ground-Truth Disparity Map**	**Result by ICBP**	**Hyp _Acc**	**Hyp _Size 1**	**Hyp _Size2**
H_Aloe				97.3	29.2	4.8
H_Baby3				97.0	41.0	4.8
H_Cloth4				98.0	31.7	4.9
H_Midd2				71.1	34.2	4.7

Dataset Name	Left Image	Ground-Truth Disparity Map	Result by ICBP	Hyp _Acc	Hyp _Size 1	Hyp _Size2
H_Rocks2				98.2	38.1	5.0
H_Wood1				97.1	29.6	4.7
L_Baby1				95.2	60.7	6.8
L_Bowling2				94.9	65.1	9.8
L_flowerpots				94.7	77.3	10.8
L_midd1				84.0	60.9	9.0
L_Monopoly				90.4	58.9	7.2
L_Wood2				95.4	58.9	9.9
H_Art				93.4	28.3	4.6
H_Books				90.2	28.7	4.6

Dataset Name	Left Image	Ground-Truth Disparity Map	Result by ICBP	Hyp _Acc	Hyp _Size 1	Hyp _Size2
H_Dolls				95.7	29.0	4.6
H_Cones				94.9	29.4	3.6
H_Loundry				88.3	28.6	4.8
H_Moeibuis				93.1	29.2	4.6
H_Reindeer				96.0	31.7	4.8
L_Teddy				95.0	61.7	7.4
H_Adirondack				82.4	26.2	2.2
H_Motorcycle				92.3	26.6	2.2
H_Pipes				92.5	23.5	2.2
H_Shelves				76.6	31.8	2.2

Dataset Name	Left Image	Ground-Truth Disparity Map	Result by ICBP	Hyp _Acc	Hyp _Size 1	Hyp _Size2
L_Jadeplant				91.0	40.9	9.0
L_MotorE				98.4	58.6	5.1
L_Piano				96.5	74.2	5.5
L_Playroom				94.7	60.6	6.5
L_Playtable				95.3	59.0	5.8
L_Vintage				90.6	37.7	8.5

The proposed algorithm for hypothesis generation based on intrinsic curves gained an average accuracy of 93±6%. That is for 93% of pixels, the ground-truth disparity value was among the hypothesized candidates for that pixel. The main reason of failure at other pixels was the fact that no hypothesis was generated at all for those pixels. They were, mainly, the pixels located at texture-less or low-textured areas of the image. At these areas, the orientation at the intrinsic curve was undefined since there were neither intensity nor gradient-of-intensity changes. Therefore, these points were left with no candidate hypothesis, unless some were found at the hypothesis extension step. That is, a pixel was locally located in the texture-less area but its close surrounding area had texture variations.

In addition, the size of candidate disparity sets for each pixel was, in average, 44±16% of the ground-truth disparity range, and only 6±2% of the image width. For algorithms such as hierarchical belief propagation (Felzenszwalb and Huttenlocher, 2006) and SGM (Hirschmüller, 2008) whose time complexity is linear with regard to the size of disparity search space, the proposed algorithm (for disparity search space reduction) can improve the speed of matching by 56% if the approximate disparity range is known, and by 94% if no information is available beforehand. Note that this reduction

is done using the original-resolution images, requiring no hierarchical search. It should also be noted that the complexity that is added to the matching algorithm for generating intrinsic curves is linear only with regard to the number of pixels in the image, and has no dependency on the disparity search range. More specifically, generating intrinsic curves requires calculating one-directional image gradients and up-sampling it at half-pixel by piecewise cubic spline interpolation. Many other robust matching costs such as the one proposed by Birchfield and Tomasi (1998), which is widely used in the literature, also require up-sampling by linear or cubic interpolations at half-pixel.

Table 6.6.4 shows the average of comparative results of the techniques mentioned at Table 6.6.2 applied to data sets of Table 6.6.3. **The detailed tables for all the datasets are provided in Appendix V.** Note that one-way results of two techniques, RSGM and SURE, are missing from Table 6.6.4 since their open-access libraries do not provide raw, one-way disparity maps. It should also be mentioned that in the experiments of this study, disparity refinement was not performed, e.g. speckle filtering, median filtering, sub-pixel refinement and hole-filling using interpolation techniques; because the objective of these experiments was to check the raw accuracy achievable only by the matching algorithms. SURE was an exception to this rule, where some post-processes such as sub-pixel refinement, filtering and hole filling were inevitable given its open-access libraries.

Table 6.6.4. Comparative results of dense matching techniques on close-range stereo images

		ICBP	ICBP _Noocc	ICSG	ICSG _Noocc	ICWTA	ICWTA _Noocc	CenWTA	RSGM	ADGC	AD HierBP	SURE
Low-resolution images	1-way error%	17.0	18.7	15.6	17.4	36.5	38.4	40.1	N/A	35.3	48.6	N/A
	2-way error%	8.8	10.8	8.8	10.8	19.6	21.7	25.4	17.9	26.5	35.1	9.3
	Density%	85.6	86.0	87.5	87.6	64.0	64.2	67.9	118.1	74.5	63.3	89.7
High-resolution images	1-way error%	22.2	23.7	19.3	21.0	45.4	47.3	47.3	N/A	40.0	37.6	N/A
	2-way error%	10.1	12.0	7.5	9.6	29.3	31.4	31.6	18.5	27.3	22.9	10.1
	Density%	78.2	78.4	79.7	79.6	55.0	55.1	55.7	119.8	63.4	69.6	86.4
Average	1-way error%	19.8	21.3	17.5	19.3	41.2	43.1	44.0	N/A	38.2	42.8	N/A
	2-way error%	9.5	11.4	8.1	10.1	24.8	26.9	28.7	18.2	27.0	28.6	9.7
	Density%	81.8	82.1	83.5	83.5	59.4	59.5	61.7	119.0	67.8	66.5	88.0

Regarding the proposed method to detect potential occlusions, a test was performed on the stereo images. For each pair, the function Θ, as defined in Equation (6), was calculated at non-occluded ground-truth pixels. Figure 6.6.1 shows the value of this function for the images of L_Dolls (note that the Θ values were ordered ascending). The similar results were obtained for other images. In average for 92±1% of the non-occluded pixels, the value of Θ was exactly zero. Therefore, the consideration of this function to estimate the occlusion term of the energy function (E_{occ}) as a soft prior (not a hard/binary constraint) was logical. The experiments showed that including the occlusion term in the energy function increased the accuracy by approximately 2%, which is not large but can avoid a considerable number of mismatches in high-resolution images.

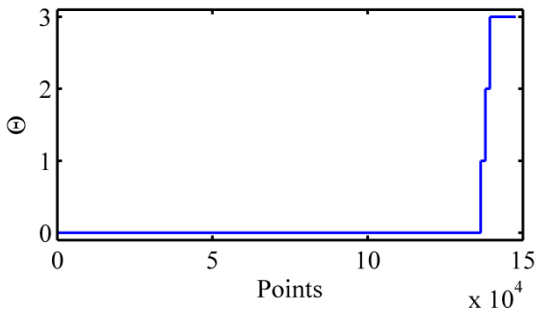


Figure 6.6.1. Values of curvature similarity, Θ, at non-occluded pixels of a stereo pair

One of the advantages of the proposed algorithm (ICSG and ICBP) over other methods was the noticeably high accuracy that could be achieved in raw one-way matching. Figure 6.6.2 shows two examples of the one-way matching results obtained by different methods. In these figures, black pixels represent an error (wrong estimation of disparity value), and white pixels represent correct matches. The percentage of bad pixels using the proposed technique was almost two times less than other techniques (difference of 20%). This feature can be beneficial to real-time mapping applications, where the speed of matching is more important than gaining metric accuracy. Therefore, without needing to perform matching on both stereo pairs, high accuracy (averagely 82%) can be achieved. Specifically, if local matching was supposed to be used, then the proposed algorithm (ICWTA) results in averagely 4% less error than a conventional local matching (CenWTA).

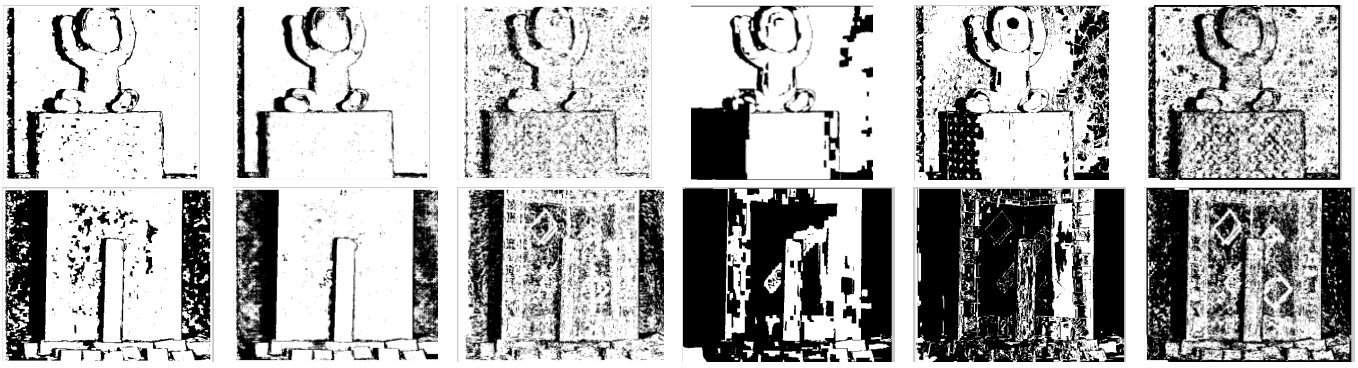
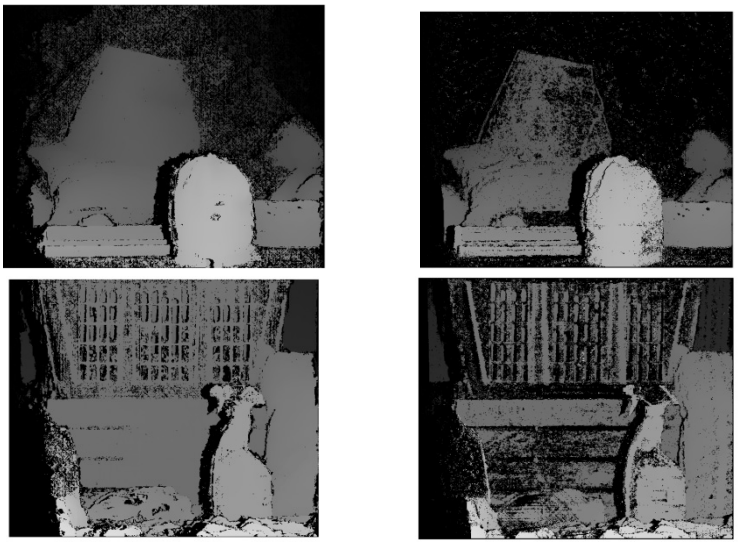


Figure 6.6.2. Examples of error maps at one-way matching by (a) ICBP; (b) ICSG; (c) ICWTA; (d) ADHierBP; (e) ADGC; (f) CenWTA

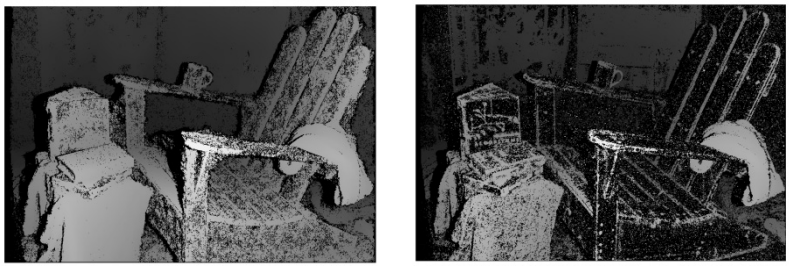
In terms of the energy minimization technique, ICSG had slightly better performance than ICBP. That is semi-globally minimizing the proposed energy function resulted in 2% higher accuracy than globally estimating the inference using LBP. In addition, the speed of matching was faster with SGM than a normal LBP, since the complexity of LBP was quadratic in terms of disparity search space and linear in terms of the number of iterations. To further test the ICSG algorithm, it was applied to 2014 full-resolution datasets of MiddleBurry benchmark. The evaluation results showed that ICSG produced the

most accurate disparity maps at badpixel=0.5 (i.e. the percentage of pixels whose estimated disparity values were 0.5 pixels or more different from the ground-truth)[1]. The average error at training and test datasets was 10% and 7% respectively. It should be noted that the results submitted to the benchmark were also raw results of dense matching with no post-processing and densification. Therefore, the evaluations of the benchmark were only valid at what is called “sparse” mode. In terms of speed, the proposed technique (ICSG) was in the middle of benchmark table; i.e. it was neither the fastest nor the slowest one. It should, however, be noted that most of the techniques of the table were implemented using CUDA or GPU, and extensive multi-core parallel instructions were applied. However, ICSG was implemented on CPU C++ and parallelizing was only partially applied at the level of matching cost calculation using parallel for-loop instructions of MSDN.

Based on Table 6.6.4, the results of SURE were more similar to ICSG than other techniques. In average, ICSG was only 1% more accurate than SURE. In average, SURE produced 9% denser disparity maps compared to ICSG, with a maximum difference of 20% in H_Loundary dataset (Figure 6.6.3). High difference of density was mainly observed in datasets like H_Midd2, H_Books and H_Adirondack. The main source of this density difference was the hypothesis-generation step at ICSG. For instance at H_Loundary dataset, only for 88% of pixels, a true disparity value was hypothesized. As explained earlier, the reason was that, for pixels located at low-textured areas of the image, such as a completely white wall, no disparity hypothesis was generated; i.e. no matches could be estimated. This, however, was not always a drawback of ICSG. In general, texture-less areas cause high matching ambiguity for any technique of matching. Therefore, the results of matching at such areas are usually not reliable. That is why ICSG had higher matching accuracy in datasets such as H_Loundary and H_Midd2 compared to SURE, since it avoided ambiguous matching.



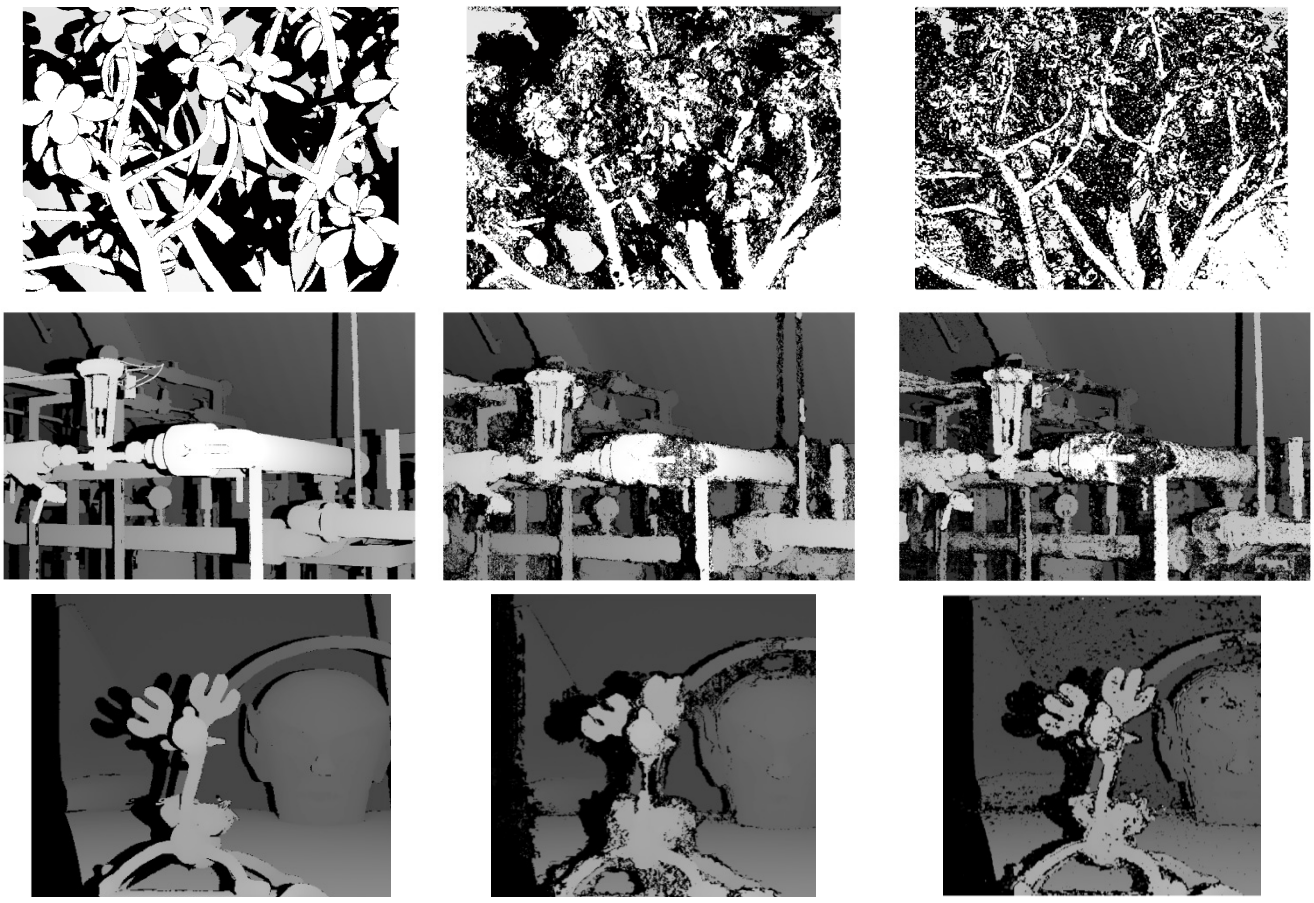
[1] Results are accessible at http://vision.middlebury.edu/stereo/eval3/ (accessed at 15/04/2016)

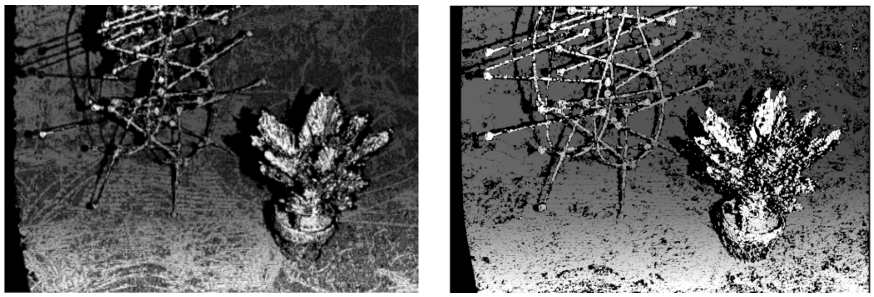


a b

Figure 6.6.3. Examples of disparity maps generated by (a) SURE and (b) ICSG. Notice their considerable differences in terms of matching density.

Another important difference was that SURE algorithm tended to ignore small details, which made it less suitable for close-range imagery. Edges were generally reconstructed more clearly using our algorithm (Figure 6.6.4). This was mainly because of the fact that in the hierarchical matching strategy of SURE, the disparity priors from low-resolution imagery were used to derive a region of interest representing the potential disparity map of the objects in high-resolution levels. Small details, like short edges, would not be present in any of the levels of the pyramid except for the original resolution. Therefore, the right disparity range of such objects might not be identified correctly, which would result in their mismatching.



N/A 

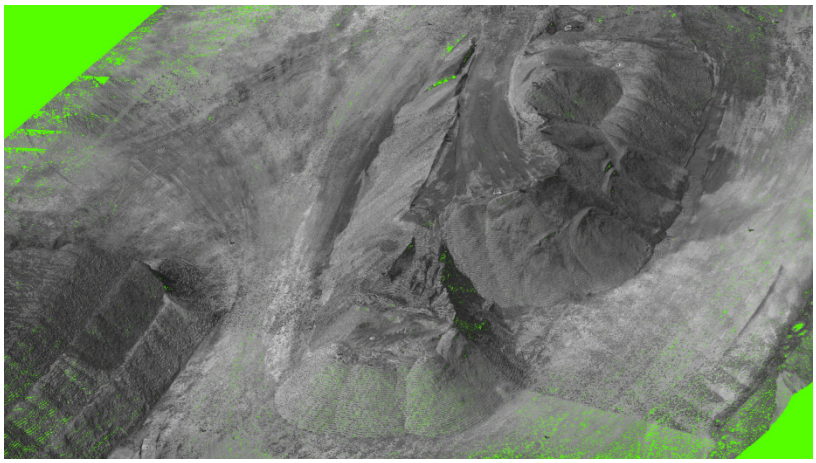
a b c

Figure 6.6.4. Examples of disparity maps and their considerable differences in terms of detail preservation. (a) Zoomed ground-truth; (b) SURE disparity map; (c) ICSG disparity map

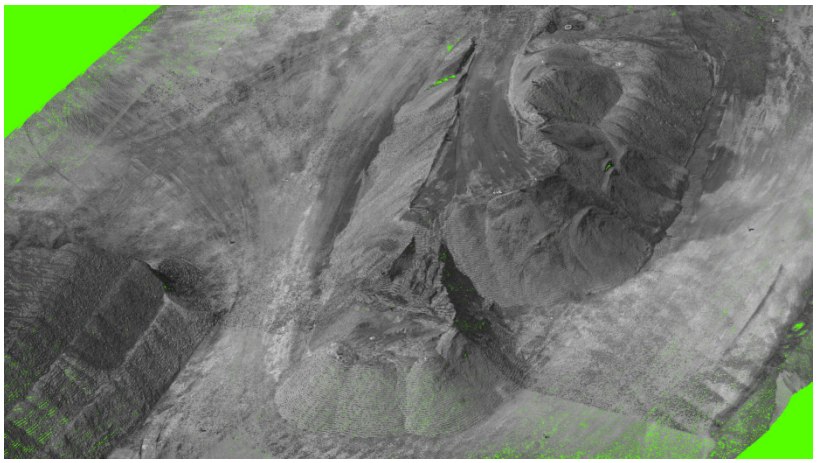
6.6.2 Aerial stereo images

The proposed algorithm was applied to unmanned aerial images captured with a high-resolution camera (sensor size of 4872×3248 pixels and focal length of 35 mm). The images were gathered from a pile at a gravel-pit mine. The dataset included 25 images that were previously epipolar rectified. So, this experiment involved only with the dense matching. As before, not post-processing was performed after either dense matching or triangulation. The 3D point clouds generated from the disparity maps of this dataset were compared with the dense 3D point cloud acquired with a LiDAR instrument (FARO Focus 3D by FARO, Lake Mary, FL, USA). The results of ICBP and ICSG were compared with that of SURE as well.

Figure 6.6.5 presents the 3D point clouds generated by each algorithm. As it can be noticed, SURE still yielded a denser result (20% denser). This difference was more evident in low-textured areas like the ground-level surface.



a



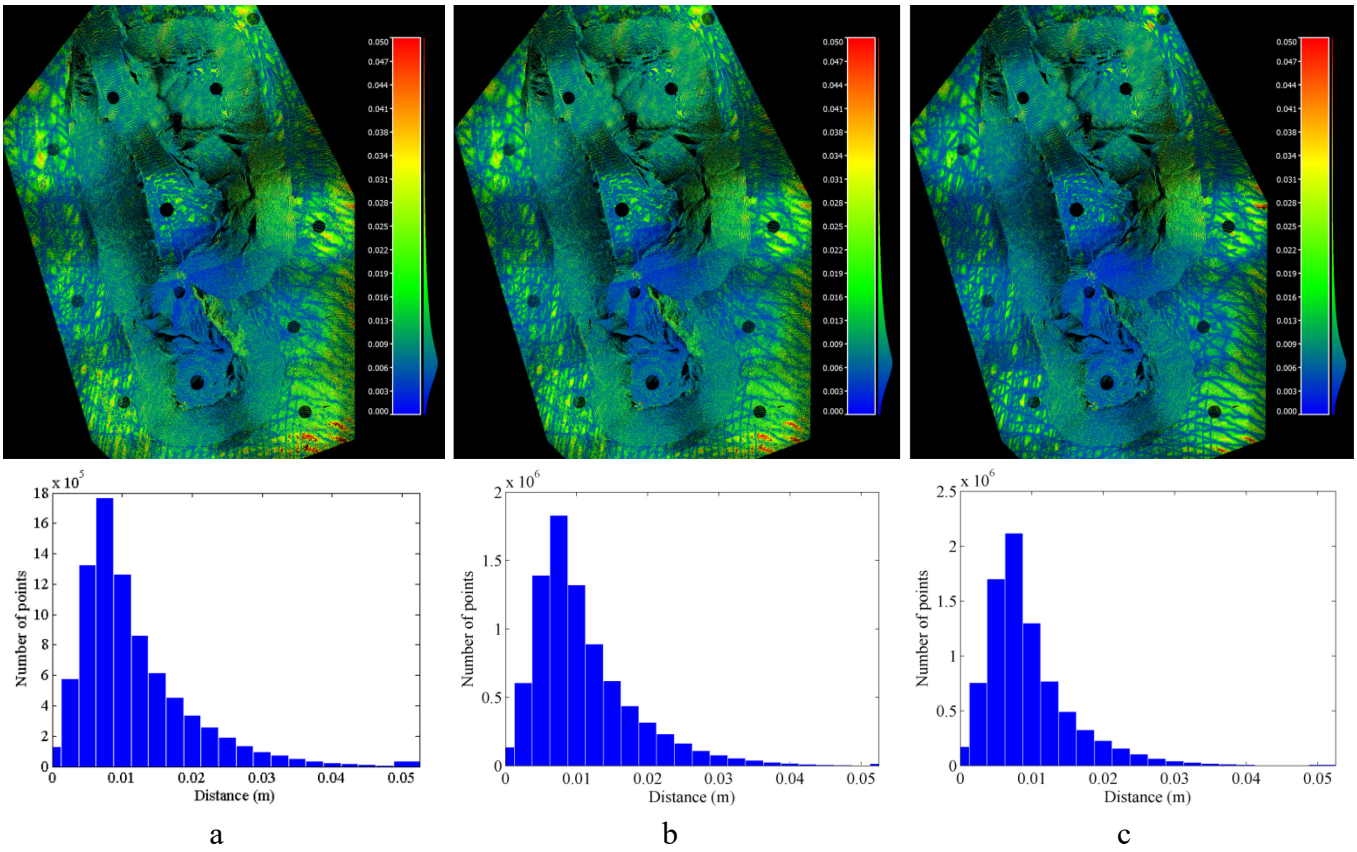
b

c

Figure 6.6.5. Point clouds generated by dense matching via (a) ICBP; (b) ICSG; (c) SURE

To evaluate the accuracy of matching, the distances of the LiDAR point cloud from the image-based point clouds were measured (Figure 6.6.6). Table 6.6.5 summarizes the histograms of distances at Figure 6.6.6.

Highest errors happened at the ground-level surface and access way of the pile. Since LiDAR acquisition was performed after image acquisition, and fieldwork team was waking at those areas, such errors were naturally expected. In a comparative sense, SURE achieved slightly better accuracy than ICSG, and ICSG performed slightly better than ICBP. However, in average, all the algorithms resulted in very high accuracy better than 12±8 mm.



a b c

Figure 6.6.6. Distance of the LiDAR point cloud from the image-based point clouds and their corresponding histograms generated by (a) ICBP dense matching; (b) ICSG dense matching; (c) SURE dense matching

Table 6.6.5. Results of comparing LiDAR point cloud to the image-based point clouds

Matching Technique	Average error (mm)	Standard deviation (mm)	RMSE (mm)
ICBP	11.5	7.7	13.8
ICSG	10.9	7.0	9.0
SURE	9.5	6.1	8.0

6.7 Conclusions

In this study, the concepts of intrinsic curves were revisited and their characteristics were studied in order to propose an efficient dense stereo matching algorithm. The first objective of this study was to reduce the disparity search range with a simple search at the space of the intrinsic curves, directly at the original resolution, and without requiring computationally complex techniques such as hierarchical matching. The suggested hypothesis generation technique considered both spatial and photometric characteristics of pixels to propose the candidate disparities. The proposed technique achieved an average accuracy of 93±6% in hypothesis generation. It was noticed that for pixels located at texture-

less (or low-textured) areas of images, no hypothesis was generated. While it was a drawback causing low-density matching, this characteristic avoided high-ambiguity matching. In addition, the size of the disparity search space was reduced to as low as 4% of the original search range. This means 96% speed achievement for most techniques of matching, whose complexity is linear with respect to the disparity search range. The second objective was to reduce the errors due to occluded pixels of images by integrating the occlusion-detection clues explicitly into the global energy function as a soft prior. These clues were also derived using the concepts of intrinsic curves and their capacity to manifest occlusions. Respectively, the inclusion of occlusion priors in the energy term increased the accuracy of matching averagely 2%. The third objective was to combine these concepts into global and semi-global matching procedure to increase the raw accuracy of dense matching. Two techniques of LBP and SGM were tested. In average, semi-global matching based on intrinsic curves gained higher accuracy. Compared to the state-of-the-art techniques, the proposed dense matching algorithm achieved considerably higher accuracy (up to 20% higher) and manifested capacity to preserve details of disparity maps. The evaluations of the proposed technique at MiddleBurry stereo benchmark also proved that it gained the most accurate results for dense matching (with no post processing).

6.8 References

Birchfield, S. and Tomasi, C., 1998. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE T Pattern Anal*, 20(4), pp. 401-406.

Bleyer, M., Rhemann, C. and Rother, C., 2011. PatchMatch stereo-stereo matching with slanted support windows. In *Proceedings of the British Machine Vision Conference*, pp. 1-11.

Boykov, Y., Veksler, O. and Zabih, R., 1998. Markov random fields with efficient approximations. In: Proceedings of IEEE Conference on Computer vision and pattern recognition, pp. 648-655.

Boykov, Y., Veksler, O. and Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE T Pattern Anal*, 23(11), Edinburgh, pp.1222-1239.

Cox, I.J., Hingorani, S.L., Rao, S.B. and Maggs, B.M., 1996. A maximum likelihood stereo algorithm. *Comput Vis Image Und*, 63(3), pp.542-567.

Felzenszwalb, P.F. and Huttenlocher, D.P., 2006. Efficient belief propagation for early vision. *Int J Comput Vision*, 70(1), pp. 41-54.

Geiger, A., Roser, M. and Urtasun, R., 2011. Efficient large-scale stereo matching. *Lecture Notes in Computer Science*, 6492, pp. 25-38.

Gurbuz, Y.Z., Alatan, A.A. and Cigla, C., 2015. Sparse recursive cost aggregation towards O(1) complexity local stereo matching. In: *Proceedings of IEEE Conference on Signal Processing and Communications Applications*, Malatya, pp. 2290-2293.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE T Pattern Anal*, 30(2), pp.328-341.

Hirschmüller, H. and Scharstein, D., 2007. Evaluation of cost functions for stereo matching. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, pp. 1-8.

Hosni, A., Bleyer, M., Gelautz, M. and Rhemann, C., 2009. Local stereo matching using geodesic support weights. In: *Proceedings of IEEE International Conference on Image Processing*, Cairo, pp. 2093-2096.

Hosni, A., Rhemann, C., Bleyer, M., Rother, C. and Gelautz, M., 2013. Fast cost-volume filtering for visual correspondence and beyond. *IEEE T Pattern Anal*, 35(2), pp. 504-511.

Kolmogorov, V. and Zabih, R., 2001. Computing visual correspondence with occlusions using graph cuts. In: *IEEE proceedings of International Conference on Computer Vision*, Vancouver, pp. 508-515.

Ohta, Y. and Kanade, T., 1985. Stereo by intra-and inter-scanline search using dynamic programming. *IEEE T Pattern Anal*, 2, pp.139-154.

Mozerov, M.G. and van de Weijer, J., 2015. Accurate stereo matching by two-step energy minimization. *IEEE T Image Process*, 24(3), pp.1153-1163.

Rothermel, M., Wenzel, K., Fritsch, D. and Haala, N., 2012. SURE: Photogrammetric surface reconstruction from imagery. In: *Proceedings of LC3D Workshop*, Berlin, pp. 1-9.

Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nesic, N., Wang, X. and P. Westling, 2014. High-resolution stereo datasets with subpixel-accurate ground truth. In: *Proceedings of German Conference on Pattern Recognition*, Münster, pp. 1-12.

Scharstein, D. and Pal, C., 2007. Learning conditional random fields for stereo. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, pp. 1-8.

Scharstein, D. and Szeliski, R., 2003. High-accuracy stereo depth maps using structured light. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, WI, pp.195-202.

Sinha, S.N., Scharstein, D. and Szeliski, R., 2014. Efficient high-resolution stereo matching using local plane sweeps. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, pp. 1582-1589.

Spangenberg, R., Langner, T., Adfeldt, S. and Rojas, R., 2014. Large scale semi-global matching on the CPU. In: *Proceedings of IEEE Symposium on Intelligent Vehicles Proceedings*, Dearborn, pp. 195-201.

Sun, J., Zheng, N.N. and Shum, H.Y., 2003. Stereo matching using belief propagation. *IEEE T Pattern Anal*, 25(7), pp.787-800.

Sun, J., Li, Y., Kang, S.B. and Shum, H.Y., 2005. Symmetric stereo matching for occlusion handling. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, pp. 399-406.

Tomasi, C. and Manduchi, R., 1998. Stereo matching as a nearest-neighbor problem. *IEEE T Pattern Anal*, 20(3), pp. 333-340.

Yoon, K.J. and Kweon, I.-S., 2006. Adaptive Support-Weight Approach for Correspondence Search, *IEEE T Pattern Anal*, 28(4), pp. 650-656.

Veksler, O., 2006. Reducing search space for stereo correspondence with graph cuts. In *Proceedings of the British Machine Vision Conference*, pp. 73.1-73.10.

Zabih, R. and Woodfill, J., 1994. Non-parametric local transforms for computing visual correspondence. *Lecture Notes in Computer Science*, 801, pp. 151–158.

Zitnick, C.L. and Kanade, T., 2000. A cooperative algorithm for stereo matching and occlusion detection. *IEEE T Pattern Anal*, 22(7), pp.675-684.

Chapter 7

7. General Conclusions and Discussions

7.1 Summary and Conclusions

The rapid advancement of UAVs and digital imaging sensors, as well as their facilitated accessibility to researchers and industries, has recently caused revolutionary movements in fields of small-scale remote sensing. As the review study of this thesis showed, the applications of these systems have grown fast and vast specifically in agricultural and natural environments. However, the speed of such growth has been so fast that researchers could not concentrate on specific aspects of these systems. In early applications, including the majority of the literature covered in the review study, in-house customized systems were more popular probably because the commercial development of UAV-remote-sensing systems for civilian use was not advanced enough. Although promising signs of success were shown by those studies, their primary shortcoming was underestimating the importance of automated data processing. UAV systems are capable of producing a large volume of data, which can end up being useless if the data-processing chain is not well adapted to it. Nowadays, among remote-sensing researchers, less attention is being concentrated on system development, since pre-packaged UAV-systems are available at wide ranges of capacities. Therefore, the current research not only should improve the data-processing schemes but also should investigate the systems extensively and realize solutions to maximize their capacities, in order to obtain satisfying results with respect to the requirements of specific applications.

In this study, the focus was on photogrammetry systems, which have had a special place in this revolutionary remote-sensing movement created by UAVs. The majority of the applications involved with unmanned aerial imagery require not only visual information of the environment but also topographic and spatial information as well. This is mainly what a photogrammetry system is about: generating three-dimensional information of the environment from the spectral data. The main question to be answered by this research was: What should be done that a UAV-PS be able to meet the 3D-modeling requirements of various engineering and natural-resource-management applications, in terms of accuracy, spatial resolution, and temporal operability? To answer this question, first, the requirements of such applications were investigated through a review study. Their issues, challenges, and future perspectives were identified. In addition, the systems used at these applications were studied, which allowed recognizing the specifications that a high-performance system should have.

Built on the lessons learned from the review study, a specific objective of this thesis was exploring the methodological and experimental aspects of developing a UAV-PS for ensuring high-quality visual and topographic data. The hardware of the system consisted of an electric-powered helicopter, a high-resolution digital camera, and an inertial navigation system. The developed system was extensively evaluated for precise modeling of an open-pit gravel mine and performing volumetric-change

measurements. This application was selected for several reasons. Firstly, from a general perspective, this application could be aligned with many other natural-resource-management applications. Secondly, this case study provided a challenging environment for 3D modeling, in terms of scale changes, terrain relief variations as well as structure and texture diversities. Thirdly, geological monitoring of mines demands high levels of accuracy, similar to the accuracy required for precision engineering applications. The third chapter of the thesis investigated issues that needed to be addressed at the levels of system development and deployment as well as data acquisition and pre-processing for producing high-quality data. The concerns regarding the accuracy of intrinsic camera calibration parameters were investigated and solutions were proposed for achieving higher accuracy at either offline calibration or on-the-job self-calibration. Factors affecting the quality of direct geo-referencing were assessed too. Regarding the system integration and multi-sensor control, it was verified that the developed software package was capable of synchronizing the navigation and imaging sensors with an approximate delay of 11 ms without requiring any additional mechanism. This is an important factor for off-the-shelf cameras, which are not necessarily equipped with a multi-sensor synchronization mechanism, such as external trigger inputs. It was also shown that platform calibration was an important factor in determining the quality of direct geo-referencing. Furthermore, several experiments were performed to assess different aspects of indirect geo-referencing. It was shown that a minimum number of GCPs could provide a high accuracy level if they were distributed evenly over the whole zone and their visibilities in images were maximized. However, under such conditions, the scale consistency of the imaging network needed to be ensured e.g. by providing proper overlap among images. That was the main reason that necessitated the planning of flight trajectory and camera shots with respect to terrain characteristics before data-acquisition. The planning software developed in this study could effectively address this point. It was also shown that the accuracy of target detection could affect the accuracy of reconstruction largely (up to 81%). Therefore, the target-detection method proposed in Chapter 3 could be used to precisely locate the GCPs on the images. This would prevent the image-measurement noise, caused by manual target detection, from affecting the accuracy of indirect geo-referencing. Moreover, the impacts of high photometric variations among images, mainly caused by shadows, on the accuracy of 3D modeling were verified. It was shown that enhancing images in a pre-processing stage not only could improve the accuracy of 3D modeling but also was mandatory in serious cases to enable the 3D reconstruction. As a result of these efforts, an operational UAV-photogrammetry system was achieved. It was demonstrated how the capacities of the system, with its current equipment and characteristics, could be maximized in different mapping circumstances. In terms of the expenses of data-acquisition with the UAV-PS compared to a manned helicopter, the most part of the economy is related to the costs of piloting services. Renting a manned helicopter for data acquisition in Quebec would cost approximately $1,500/hour[1], while the piloting service of the UAV used in this study cost only $150/hour. In addition, UAVs, that are more recent[2], do not require specific piloting skills, which eliminate this expense as well. In terms of equipment, the rental fees for the sensors and land-surveying equipment required for data acquisition were as follows. The Responder platform, Prosilica GE4900C camera, and Microbotics INS would approximately cost

[1] The prices mentioned in the thesis are based the rates used by Centre de geomatique du Quebec at the time this project was conducted.

[2] For instance, DJI platforms (www.dji.com)

$450/day[1]. The Trimble R8 GNSS mobile and base receivers with accessories would approximately cost $350/day. If the data-acquisition sessions were well planned, then few hours of fieldwork would be sufficient. Therefore, the daily cost of using the UAV-system developed in this study would be less than $1,850. It should be noted that the UAV technology has advanced with an incredible pace between the time this project was started and the time this thesis is presented. Therefore, nowadays, less expensive sensors and platforms are available in the market for comparable prices and higher performances. For instance, Sony Alpha a7R II with almost three times the resolution of GE4900C is currently a popular camera for integration to UAVs. Besides, specialized navigation sensors are developed for UAVs, such as Trimble APX-15 UAV, which has RTK capability and can provide very accurate direct geo-referencing of images.

The other specific objective of this thesis was developing photogrammetric-processing techniques that were adapted to the characteristics of low-altitude unmanned imagery. Firstly, a modified version of the integer-coded genetic algorithm for the problem of robust epipolar-geometry estimation from putative matches was proposed. It was followed by an adaptive thresholding algorithm for inlier classification. The main issue that this algorithm needed to address was the effective sparse matching of stereo images majorly contaminated by noise and outliers. Extensive experiments were performed to assess the components of this algorithm and its overall performance in comparison with the state-of-the-art techniques (mainly RANSAC-like methods). The proposed approach showed considerable robustness to high percentages of outliers, to degenerate configurations and to image noise. It was shown that the computational expenses of the algorithm were not increasing with the ratio of outliers (even when growing over 80%) or the magnitude of image noise. These characteristics were mostly due to the guided sampling strategy, the evolutionary search scheme and inlier classification based on adaptive thresholding. Secondly, a BBA strategy for on-the-job self-calibration based on the integration of pseudo-observations to Gauss-Helmert model was assessed. It was shown that this simple strategy could significantly decrease the correlation of IO and EO parameters in the bundle adjustment (60% improvement), without requiring any additional orientation observations. As a result, the accuracy of reconstruction could be improved over 80%. It was also shown that simultaneous adjustment of observations with unknowns in the GHM could result in extra 12% of accuracy improvement. Thirdly, the concepts of intrinsic curves were revisited in order to propose an efficient dense stereo matching algorithm. To initiate dense matching, a hypothesis-generation technique was proposed, which considered both spatial and photometric characteristics of pixels to propose the candidate disparities for them. It was shown that the proposed solution was able to reduce the disparity search range with a simple search at the space of the intrinsic curves, directly at the original resolution, and without requiring computationally complex techniques such as hierarchical matching. The proposed technique achieved an average accuracy of 93±6% in hypothesis generation, and could reduce the size of the disparity search space to as low as 4% of the original search range. In addition, this technique was avoiding ambiguous matching at texture-less areas of images. In order to reduce the errors due to

[1] This rental fees are based on the approximate original price of the equipment. Note that the equipment was acquired several years ago at considerably high prices. Nowadays, the same performance can be expected from equipment available at the market with lower prices.

occluded pixels, specific occlusion clues were derived using the curvature characteristics of intrinsic curves. They were explicitly integrated into the global energy function as a soft prior. Including these occlusion priors in the energy term increased the accuracy of matching. Compared to the state-of-the-art, the proposed dense matching algorithm achieved considerably higher accuracy (up to 20% higher). The evaluations of the proposed technique at MiddleBurry stereo benchmark also proved that it gained the most accurate results for dense matching (with no post processing).

Finally, it was shown that the application of the proposed photogrammetric techniques to the imagery acquired by the developed UAV-PS resulted in average dense reconstruction accuracy of 11±7 mm, and sparse reconstruction accuracy of 6±1 mm at spatial resolution of better than 1.3 cm. Such high level of accuracy and resolution can meet the requirements of many geological, natural-resource-management and precise engineering applications. For instance, the topographic data required for predicting hazardous events, such as ground subsidence, slope instability and landslides, must provide a ground resolution of 1-3 cm. The map scale required for volumetric measurement in earthworks is usually between 1:4000 and 1:10,000, which is roughly equivalent to 3-7 cm of imaging spatial resolution with GE4900C camera. Therefore, in these cases, the system is operational and the level of accuracy achievable with the system is sufficient. However, the spatial resolution, thus the mapping accuracy, might need to be improved for applications that require higher precision, such as detecting structural fractures with a few millimeters width and inspecting displacements of infrastructures limited to a few millimeters. The success of the proposed algorithms on both close-range imagery and the imagery taken from the gravel-pit mine would mean that they could be applicable to other types of UAV-imagery as well, e.g. high-oblique imagery required for inspecting vertical structures such as building facades.

The main contributions and originalities of this thesis can be summarized to the following ones.

1. Critical review of remote-sensing applications of unmanned aerial imagery
2. Comprehensive theoretical and experimental study of the requirements of a high-performance UAV-photogrammetry system at levels of implementation, deployment and data acquisition, and proposing solutions for the common challenges existing in the current literature with these regards
3. Novel and effective techniques of sparse and dense reconstruction
    3.1. Developing a new technique of robust epipolar-geometry estimation and inlier detection for sparse matching, which showed considerable robustness (while maintaining high efficiency) against high outlier ratios, noise in observations and degenerate configurations compared to the state-of-the-art
    3.2. Suggesting a BBA strategy and its sparse implementation for de-correlating IO and EO parameters in bundle adjustment of a large number of images
    3.3. Proposing a new technique of dense stereo matching by revisiting the concepts of intrinsic curves that showed high efficiency in handling large disparity search ranges, occlusions, and matching ambiguities

7.2 Research Perspectives and Future Work

In the future, the performance of the sparse- and dense-reconstruction techniques will be optimized. More specifically, the genetic algorithm will be implemented in parallel way to increase the speed of the algorithm. The strategy of the guided sampling will be improved in order to take the quality of correspondences into account, for instance the results of detecting of noise, foreground motion, and non-rigid structures. For the algorithm of dense matching, the density of matching at texture-less areas should be improved. This will be done by integrating shape/structure priors into the global energy functions. The raw results of dense matching will also be post-processed for increasing the accuracy.

The achievements of this thesis proved that obtaining very high levels of 3D-modelling accuracy with UAV-photogrammetry systems is possible. However, accuracy is only one of the criteria to measure the capacity of UAV-PSs; autonomy and intelligence are other important factors. Therefore, we are interested in investigating UAV-PSs for applications were all these three factors are important, e.g. metric inspection of vertical infrastructures and post-disaster safety inspection in urban canyons. These applications are quite challenging mainly because of the limited knowledge available beforehand about their environment. Hence, the trajectory designed for the UAV flight will not be sufficiently reliable. As a result, significant pilot effort should be continually focused on keeping camera orientation over the features of interest and on the safety of the vehicle with respect to the unknown obstacles in the environment. Moreover, GNSS-based navigation will be a challenging task in such environments specifically due to GNSS-signal blockage and communication loss with the base stations. For these reasons, there will be a considerable area for improving the efficiency of flight through the development of better path-planning strategies and navigation techniques. We are specifically interested in improving the UAV-PS developed at this study by adding a multi-modal navigation system to it through integrating GNSS-based inertial sensors with vision-based, laser-based and wireless positioning systems. Such a system will be able to perform both accurate trajectory-following and online decision-making and mission planning (e.g. sense and avoid).

Another element that makes these applications challenging is that their end-users do usually not have any expertise in working with three-dimensional data to extract the specific types of information they require rapidly and efficiently. Therefore, such applications demand automated semantic image processing and intelligent scene understanding. For instance, metric inspection of buildings' facades and roofs would require a 3D line cloud (instead of a point cloud) representing the building outlines along with segments representing different sorts of anomaly (e.g. thermal and structural ones). We are particularly interested in extending our sparse- and dense-reconstruction algorithms with integrating semantic properties of the objects directly into them, and using this to meet the requirements of high-level engineering-grade applications.

8. References

Abdel-Aziz, Y.-I., and H. Karara (1971) Direct linear transformation from comparator coordinates in close-range photogrammetry, In Proceedings of the ASP Symposium on Close-Range Photogrammetry, Urbana, Illinois, p. 1–18.

Ackermann, F. (1981) Block adjustment with additional parameters. Photogrammetria vol. 36, n^{o} 6, p. 217-227.

Agarwal, S., Snavely, N., Simon, I., Seitz, S.-M. and Szeliski, R. (2009) Building rome in a day. In Proceedings of IEEE International Conference on Computer Vision, Kyoto, p. 72-79.

Agarwal, S., Snavely, N., Seitz, S.-M. and Szeliski, R. (2010) Bundle adjustment in the large. Lecture Notes in Computer Science, vol. 6312, p. 29-42.

Apollonio, F.-I., Ballabeni, A., Gaiani, M. and Remondino, F. (2014) Evaluation of feature-based methods for automated network orientation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XL-5, p. 47-54.

Byröd, M. and Åström, K. (2009) Bundle adjustment using conjugate gradients with multiscale preconditioning. In Proceedings of British Machine Vision Conference. London, UK, p. 1-10.

Engel, J., Schöps, T. and Cremers, D. (2014) LSD-SLAM: Large-scale direct monocular SLAM. Lecture Notes in Computer Science, vol. 8690, p. 834-849.

Engels, C., Stewénius, H. and Nistér, D. (2006) Bundle Adjustment Rules. Photogrammetric Computer Vision, vol. 2, p. 124-131.

Fusiello, A., Trucco, E. and Verri, A. (2000) A compact algorithm for rectification of stereo pairs. Machine Vision and Applications, vol. 12, n^{o} 1, p. 16-22.

Granshaw, S.I. and Fraser, S.C., 2015. Computer vision and photogrammetry: interaction or introspection? The Photogrammetric Record, vol. 30, n^{o} 149, p. 3-7.

Haala, N., Cramer, M. and Rothermel, M. (2013) Quality of 3D point clouds from highly overlapping UAV imagery. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Rostock, Germany, vol. XL- 1/W2, p. 183-188.

Hartley, R. and Zisserman, A. (2003) Multiple view geometry in computer vision. Cambridge University Press, Cambridge, UK, 657 p.

He, F. and Habib, A. (2014) Automatic orientation estimation of multiple images with respect to laser data. Presented at the ASPRS 2014 Annual Conference, Louisville, USA.

Hu, Q. and Ai, M. (2011) A scale invariant feature transform based matching approach to unmanned aerial vehicles image geo-reference with large rotation angle. in Proceedings of IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, June 2011, p. 393-396.

Indelman, V., Roberts, R., Beall, C. and Dellaert, F. (2012) Incremental light bundle adjustment. In Proceedings of the British Machine Vision Conference, Surrey, UK, p.134.1-134.11.

Kume, H., Sato, T. and Yokoya, N. (2015) Bundle adjustment using aerial images with two-stage geometric verification. Computer Vision and Image Understanding, vol. 138, p. 74-84.

Li, J. and Allinson, N.-M. (2008) A comprehensive review of current local features for computer vision. Neurocomputing, vol. 71, nº 10, p. 1771-1787.

Linder, W. (2006) Digital photogrammetry-a practical course. Springer, Netherlands, 214 p.

Lourakis, M.-I. and Argyros, A.-A. (2009) SBA: A software package for generic sparse bundle adjustment. ACM Transactions on Mathematical Software, vol. 36, no 1, p. 2:1-2:30.

Lowe, D.-G. (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, vol. 60, nº 2, p. 91-110.

Luhmann, T., Robson, S., Kyle, S. and Harley, I. (2007) Close Range Photogrammetry: Principles, Techniques and Applications; John Wiley & Sons: Hoboken, NJ, USA, 510 p.

Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F. and Sayd, P. (2009) Generic and real-time structure from motion using local bundle adjustment. Image and Vision Computing, vol. 27, p. 1178-1193.

Mukherjee, D., Wu, Q.-J. and Wang, G. (2015) A comparative experimental study of image feature detectors and descriptors. Machine Vision and Applications, vol. 26, nº 4, p. 443-466.

Ni, K., Steedly, D. and Dellaert, F. (2007) Out-of-core bundle adjustment for large-scale 3D reconstruction. In Proceedings of IEEE International Conference on Computer Vision, Rio de Janeiro, p. 1-8.

Pang, S., Xue, J., Wang, L. and Zheng, N. (2012) Large-Scale bundle adjustment by parameter vector partition. Lecture Notes in Computer Science, vol. 7727, p. 26-39.

Pullivelli, A.M., 2006. Low-cost digital cameras: calibration, stability analysis, and applications. Master of Science thesis, University of Calgary, 170 p.

Ressl, C. (2003) Geometry, constraints and computation of the trifocal tensor. Doctoral dissertation, Technische Universität Wien, 207 p.

Triggs, B., McLauchlan, P., Hartley, R.I., Fitzgibbon, A. (1999) Bundle Adjustment – A modern synthesis. Lecture Notes in Computer Science, vol. 1883, p. 298-372.

Watson, G.-A. (2006) Computing Helmert transformations. Journal of Computational and Applied Mathematics, vol. 197, nº 2, p. 387-394.

Wells, D.-E. and Krakiwsky, E.-J. (1971) The method of least squares. Department of Geodesy and Geomatics Engineering, University of New Brunswick.

Wolf, P.-R., Dewitt, B.-A. (2000) Elements of Photogrammetry: With Applications in GIS, 3rd ed. The McGraw-Hill Companies: Boston, MA, USA, 696 p.

Wu, C., Agarwal, S., Curless, B. and Seitz, S.-M. (2011) Multicore bundle adjustment. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI , p. 3057-3064.

Zhang, Y., Xiong, J. and Hao, L. (2011) Photogrammetric processing of low-altitude images acquired by unpiloted aerial vehicles. The Photogrammetric Record, vol. 26, p. 190-211

Appendix I. Intrinsic Camera Calibration

The practical steps of off-line camera calibration and image correction are as follows.

i. Decide the parameters of the calibration model.
ii. Design and establish a test-field with signalized targets.
iii. Set the camera, pose it at different orientations, and photograph the test-field.
iv. Detect and position the targets on the images.
v. Perform self-calibration to calculate the calibration parameters.
vi. Correct the images to make them undistorted.

The design and configuration of the test field are thoroughly discussed in Section 3.4.1.1. Detection and positioning the targets is also presented in Section 3.4.1.2. The following sub-sections describe the camera calibration model and the estimation procedure.

AI.1 Camera Calibration Model

The conventional imaging system in photogrammetry is called the principal coordinate system (Figure AI.1). The center of the principal coordinate system is the camera perspective center (point *C* in Figure AI.1). The optical axis passing through the perspective center intersects the image plane at the principal point (point *p.p.* in Figure AI.1). The z-axis is aligned to the optical axis. The x-axis is parallel to the sensor horizontal edge and directed as in the conventional, pixel-based coordinate system. The y-axis keeps the coordinate system right-handed. Nominally, the principal point is located at the center of the sensor, and the sensor elements (pixels) are considered as perfect squares.

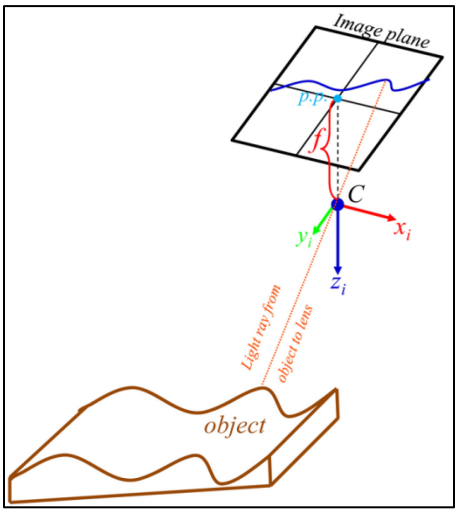


Figure AI.1. The imaging coordinate system

Now, consider a point measured at coordinates (x, y) in the nominal principal coordinate system. These coordinates must be corrected by $(\Delta x, \Delta y)$ to take the systematic errors and the principal point offsets into account. In this project, the Brown's additional parameters set for digital cameras are applied to

model the systematic errors (Dörstel *et al*., 2003; Brown, 1971; Fryer and Brown, 1986). Therefore, the main equations to determine $(\Delta x, \Delta y)$ can be stated as follows.

$$\begin{aligned}\Delta x &= -x_p + (x - x_p).(K_1 r^2 + K_2 r^4 + K_3 r^6) + P_1(r^2 + 2(x - x_p)^2) \\ &+2P_2(x - x_p).(y - y_p) + A_1(x - x_p) + A_2(y - y_p)\end{aligned} \quad (1)$$

$$\Delta y = -y_p + (y - y_p).(K_1 r^2 + K_2 r^4 + K_3 r^6) + P_2(r^2 + 2(y - y_p)^2) + 2P_1(x - x_p).(y - y_p) \quad (2)$$

In these equations, (x_p, y_p) are the principal point offsets, and r is the radial distance of the observation from the principal point, which is defined as $r = \sqrt{(x - x_p)^2 + (y - y_p)^2}$.

The error terms considered in Equation 1 and Equation 2 are described in the following paragraphs.

i) Parameters of radial $(K_1, K_2, ...)$ and tangential lens distortion (P_1, P_2)

Although lens distortions might affect the image in an irregular manner, they are mostly symmetrical in a radial sense. This effect is originated from the physical symmetry of the lens structure. Radial distortions might be observed in two different forms, namely barrel and pincushion distortions. The former causes the image to seem inflated by decreasing the image magnification as the distance from the optical axis increases. The latter makes the image look pinched at the center by increasing the image magnification with the distance from the optical axis. The complex effect, or mustache distortion, is caused by the combination of both barrel and pincushion distortions.

Tangential or decentering lens distortion is caused by misalignment of physical elements in the lens. Imagining some hypothetical circles originating from the principle point at the image plane, the decentering distortion vectors are tangent lines to these circles, while radial distortion vectors are the circles radii.

ii) Sensor, in-plane distortion parameters (A_1, A_2)

This sort of systematic error is caused by the physical properties of the sensor elements (pixels). The parameters (A_1, A_2) model this error by a first order, two-dimensional polynomial. The term A_1 is the scale parameter, which takes the non-square size of pixels into account. The term A_2 is the shear parameter; it denotes the declination of the angle between the sensor x and y directions from perpendicular status, namely non-orthogonality of the pixel array.

AI.2 Bundle Adjustment with Additional Parameters

Briefly summarizing the procedure of calibration, a test-field is first needed. The test-field includes several targets installed in fixed positions. Once the test-field is established, the camera is moved to various positions and orientations in front of the test-field, and the test-field is photographed at each position. This results in a network of images, from which the targets must be detected; that is the coordinates of all the targets should be determined on every image. The output of all these procedures

would be a set of image coordinates as (x_{ij}, y_{ij}) representing the coordinates of target number *i* in the j^{th} image.

In order to determine the intrinsic camera calibration parameters, self-calibration should be performed. The observation equations used for self-calibration are based on co-linearity condition as defined in Equation (3).

$$\begin{aligned} x_{ij} + \Delta x_{ij} + f\frac{U_{ij}}{W_{ij}} &= 0 + \varepsilon_x \\ y_{ij} + \Delta y_{ij} + f\frac{V_{ij}}{W_{ij}} &= 0 + \varepsilon_y \end{aligned} \quad . \tag{3}$$

where,

$$\begin{bmatrix} U_{ij} \\ V_{ij} \\ W_{ij} \end{bmatrix} = R(\omega_j, \varphi_j, \kappa_j) \begin{bmatrix} X_i - X_j^C \\ Y_i - Y_j^C \\ Z_i - Z_j^C \end{bmatrix}. \tag{4}$$

In general, the unknowns of these equations are

- the EO parameters $(\omega_j, \varphi_j, \kappa_j, X_j^C, X_j^C, X_j^C)$ for *j*=1 to the number of images in the network,
- the model-space coordinates of the targets (X_i, Y_i, Z_i) for *i*=1 to number of targets in the test-field, and
- the intrinsic calibration parameters $(x_p, y_p, f, K_1, K_2, K_3, P_1, P_2, A_1, A_2)$ that are involved in computing $(\Delta x_{ij}, \Delta y_{ij})$ as in Equation 1 and Equation (2).

To solve these equations and determine the unknowns, free-network bundle adjustment with additional parameters should be performed. The inner constraints (three constraints for the position, three for the orientation, and one for the scale of the model-space coordinate system) are also applied as datum constraints to eliminate the singularity of the problem.

Since the co-linearity equations (Equation (3)) are nonlinear with respect to the unknowns, initial estimations of the unknowns are required to perform the adjustment. In our case, the approximate 3D coordinates of targets were measured using a spatial station. These coordinates were also used to estimate the approximate values of EO parameters via single image space-resection. There are several linear methods in the literature for this purpose such as direct linear transformation (DLT) (Abdel-Aziz and Karara, 1971). In this study, an approximate, closed-form Perspective 3-Points (P3P) solution (or resection with minimum object information) is used since high accuracy is not required, and these estimations are only used to initiate the iterative bundle adjustment procedure. In this method, an intermediate coordinate system is used to calculate the approximate values of the rotation matrix and translation vector of the camera (Luhmann et al., 2007, p. 46).

Assuming that *xyz* represents the camera coordinate system and *XYZ* represents the object (model-space) coordinate system, the *uvw* represents the intermediate coordinate system at the scale of the camera and *UVW* at the scale of the object. Assume that there are three reference points as $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, $\boldsymbol{p}_3$ on the image with known 3D coordinates as objects $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, $\boldsymbol{P}_3$. The intermediate coordinate systems can be reconstructed as follows.

$$\begin{aligned} U &= \frac{P_2 - P_1}{|P_2 - P_1|} & u &= \frac{p_2 - p_1}{|p_2 - p_1|} \\ W &= \frac{U \times (P_3 - P_1)}{|U \times (P_3 - P_1)|} & w &= \frac{u \times (p_3 - p_1)}{|u \times (p_3 - p_1)|} \\ V &= W \times U & v &= w \times u \end{aligned} \tag{5}$$

Vector $\boldsymbol{u}$ is the unit vector on the u-axis and vector $\boldsymbol{w}$ is perpendicular to the uv-plane. These vectors can define the rotation matrices from *uvw* to *xyz* and to *XYZ* as follows.

$$\begin{aligned} R_{uvw}^{XYZ} &= [U \quad V \quad W] \\ R_{uvw}^{xyz} &= [u \quad v \quad w] \end{aligned} \tag{6}$$

Therefore, the approximate rotation matrix from *xyz* to *XYZ* can be determined as:

$$R_{xyz}^{XYZ} = R_{uvw}^{XYZ} \left(R_{uvw}^{xyz} \right)^T . \tag{7}$$

Using the centroids of the reference points in camera coordinate system ($\mathbf{x_s}$) and in the object coordinate space ($\mathbf{X_s}$), approximate values for the translation vector (T) can be defined as follows.

$$T = X_s - R_{xyz}^{XYZ} x_s . \tag{8}$$

AI.3 Stability Analysis

The intrinsic camera calibration was repeated several times, and the parameters were compared for stability analysis. From the statistical point of view, a T-test was performed. For testing the null hypothesis that the population mean of each parameter is equal to the average value of that parameter, the statistic $t = \frac{(\bar{p} - \mu_0)}{s / \sqrt{n}}$ was used, where is the sample mean $\bar{p}$, s is the sample standard deviation of the sample and n is the sample size. The degrees of freedom used in this test were n-1. Table AI.1 presents the average values, variances, and the test statistics, given a significance level of 0.05, n=17. This states that, e.g., with a confidence 0.95, the population mean of focal length would be a value between 35.96 and 36.08.

Table AI.1. Average values of intrinsic calibration parameters

Parameter	Average value	Variance	T-test variance
K_1	7.287E-05	4.592E-11	2.856E-11

K_2	-5.884E-08	1.124E-15	6.993E-16
K_3	1.390E-11	1.764E-21	1.097E-21
P_1	3.124E-06	1.275E-10	7.933E-11
P_2	5.184E-06	4.743E-10	2.950E-10
A_1	1.550E-04	2.102E-08	1.308E-08
A_2	-5.461E-05	9.344E-09	5.813E-09
x_{pp}	-1.214E-01	6.302E-03	3.920E-03
y_{pp}	1.905E-01	3.005E-02	1.870E-02
f	3.602E+01	6.077E-03	3.780E-03

According to this test, the results of calibration confirmed the null hypothesis at some dates, and the results rejected the null hypothesis at other dates. The variances of those calibration values that confirmed the null hypothesis were used in the BBA strategy proposed at Chapter 5.

In order to physically test the stability of calibration parameters, a comparison test was performed based on the similarity of reconstructed bundles using single photo resection (SPRT) method (Pullivelli, 2006 , p. 44). Table AI.2 presents the spatial offsets (σ_0 in pixels) obtained by comparing intrinsic camera calibration parameters of every two tests using the SPR method. One can notice that the IOPs of the camera calculated at close dates are similar ($\sigma_0<0.7 pixels$). However, in general, these parameters are not similar, specifically, as the camera lens was unattached and the focus distance was slightly changed in-between two tests. Therefore, it was suggested to re-estimate the intrinsic calibration parameters through on-the-job self-calibration instead of relying solely on the offline calibration parameters.

Table AI.2. Stability comparison of IOP sets using SPR method (σ_0 in pixels)

Date (m/d)	10/28	10/28	4/28	4/28	5/7	5/7	5/12	5/12	6/18	6/18	7/9	7/9	9/4	9/4	9/4	8/18
07/10	0.6	0.5	2.1	2.1	2.2	2.5	1.6	1.5	0.6	2.2	1.2	1.2	2.5	1.6	2.2	2.8
10/28	-	0.1	2.7	2.7	2.8	3.1	2.1	2.1	1.2	2.8	1.8	1.8	3.1	2.2	2.8	3.3
10/28	-	-	2.7	2.6	2.8	3.0	2.1	2.0	1.2	2.7	1.7	1.8	3.0	2.1	2.7	3.2
4/28	-	-	-	0.1	0.2	0.5	0.8	0.9	1.5	0.2	1.0	0.9	0.6	0.6	0.7	1.5
4/28	-	-	-	-	0.3	0.5	0.8	0.8	1.4	0.2	0.9	0.9	0.7	0.6	0.7	1.5
5/7	-	-	-	-	-	0.3	0.8	0.9	1.6	0.1	1.1	1.0	0.4	0.7	0.5	1.5
5/7	-	-	-	-	-	-	1.1	1.1	1.9	0.3	1.3	1.3	0.3	0.9	0.6	1.5
5/12	-	-	-	-	-	-	-	0.1	0.9	0.8	0.6	0.5	1.1	0.5	0.8	2.1
5/12	-	-	-	-	-	-	-	-	0.9	0.9	0.5	0.5	1.2	0.5	0.9	2.1
6/18	-	-	-	-	-	-	-	-	-	0.5	0.6	0.6	1.9	1.0	1.6	2.3
6/18	-	-	-	-	-	-	-	-	-	-	1.0	0.9	0.5	0.6	0.5	1.5
7/9	-	-	-	-	-	-	-	-	-	-	-	0.1	1.3	0.4	1.0	1.9
7/9	-	-	-	-	-	-	-	-	-	-	-	-	1.3	0.4	1.0	1.8
9/4	-	-	-	-	-	-	-	-	-	-	-	-	-	0.6	0.4	1.6
9/4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.6	1.6
9/4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.7

AI.4 Image Correction

Image correction means to create an undistorted image from the raw image. Consider an image point, which is observed on the pixel (u,v), where u and v are the column and row numbers. Therefore, its coordinates in the nominal principal coordinate system can be computed as $(x = u.s - W/2, -v.s + H/2)$. W and H are the width and the height of the sensor, and s is the pixel size. Assume that the gray value (intensity) observed at this point is G. Then, it can be stated that this spectral data should have been observed at another point with coordinates $(x + \Delta x, y + \Delta y)$, where $(\Delta x, \Delta y)$ are expressed as in Equation 1 and Equation 2. Therefore, the task of image correction is to map the observed spectral data to their corrected locations.

To correct the image via this concept, the regular grid of the image (the pixels on integer rows and columns) are corrected and moved to decimal positions. The gray values of these points are equal to the gray values of the initial pixels. Now, in order to produce an image array from the irregular corrected points, an interpolation technique can be used. The most common types of interpolation are the nearest neighbor, bilinear and cubic methods.

Appendix II. Initial Recovery of EO Parameters and 3D Object-Space Coordinates of Corresponding Points

AII.1. RO Parameter Estimation

The output of robust epipolar-geometry estimation is a fundamental matrix (F), which comprises the following elements.

$$F = K_2^{-\mathrm{T}} [\boldsymbol{t}]_\times R_1^2 K_1^{-1} \tag{1}$$

where K_2 and K_1 are the calibration matrices of the second and the first camera respectively. (R,$\boldsymbol{t}$) represent the rigid body transformation (rotation and translation) from the first camera coordinate system to the second one, namely RO parameters. $[\boldsymbol{t}]_\times$ is the anti-symmetric matrix defined from $\boldsymbol{t}$.

Using the offline intrinsic camera calibration parameters, or the nominal approximate intrinsic calibration parameters, matrices K_2 and K_1 can be estimated and eliminated from Equation (1). This leaves the essential matrix, E, as follows.

$$E = [\boldsymbol{t}]_\times R_1^2 \tag{2}$$

Assume that $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ are the coordinates of the perspective centers of the first and the second cameras, respectively, in the object-space coordinate system (n), R_1^n and R_2^n are the rotation matrices to make the principal coordinate systems of the first and the second cameras, respectively, parallel to n. Therefore $\boldsymbol{t}$ and R_1^2 are defined as:

$$\begin{aligned} \boldsymbol{t} &= \left(R_2^n\right)^T \left(\boldsymbol{t}_2 - \boldsymbol{t}_1\right), \\ R_1^2 &= \left(R_2^n\right)^T R_1^n. \end{aligned} \tag{3}$$

Given the essential matrix (E), there are two ways to determine the parameters of relative rotation and translation: linear way and nonlinear way. The linear way is usually used to provide the non-linear way with initial estimations. In this study, the linear solution based on singular value decomposition of the essential matrix is used (Hartley and Zisserman, 2003, p. 258). Then, the nonlinear solution based on co-planarity condition (Wolf and Dewitt, 2000, p. 236) is used to refine the RO parameters.

The essential matrix has only five degrees of freedom. The translation vector $\boldsymbol{t}$ has scale ambiguity, therefore, has only two degrees of freedom. The rotation matrix, although composed of nine elements, is an orthogonal matrix where the sum of squared elements on each row and each column is one (resulting in six constraints). Therefore, it has only three degrees of freedom. Therefore, the first and the second singular values of any essential matrix are equal, and the third one is zero. If the SVD of E is $Udiag(1,\ 1,\ 0)V^T$, then there are two possible factorizations for essential matrix $E = SR$ as:

$$S = UZU^T, \quad R = UWV^T \quad \text{or} \quad UW^TV^T \tag{4}$$

where,

$$W=\begin{bmatrix}0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix},\quad Z=\begin{bmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix} \tag{5}$$

In Equation (4), R is the same as R_1^2 and S is the same as $[\boldsymbol{t}]_\times$ up to scale. Therefore, $S\boldsymbol{t}=0$, which follows that $\boldsymbol{t} = U\,(0, 0, 1)^{\mathrm{T}} = \boldsymbol{u}_3$, the last column of U. It can be concluded that there are four following choices for the relative orientation parameters given an essential matrix.

$$\begin{aligned}&1.\ R=UWV^T\ ,\ \boldsymbol{t}=\boldsymbol{u}_3\\ &2.\ R=UWV^T\ ,\ \boldsymbol{t}=-\boldsymbol{u}_3\\ &3.\ R=UW^TV^T\ ,\ \boldsymbol{t}=\boldsymbol{u}_3\\ &4.\ R=UW^TV^T\ ,\ \boldsymbol{t}=-\boldsymbol{u}_3\end{aligned} \tag{6}$$

From these four options, only the one that results in 3D reconstruction in front of both cameras is the correct one (Hartley and Zisserman, 2003, p. 518). A point $\boldsymbol{X}$ with Homogeneous coordinates *(X,Y,Z,1)* is in front of a camera with projection matrix P if the sign of its depth from the camera, *depth(**X**;P)*, is positive. Given the case of relative orientation, the first camera can be set as the reference with camera projection matrix $P_1=(K_1)^{-1}[I_3\,|\,0]$, where I_3 is a 3×3 identity matrix. Respectively, the camera matrix of the second camera can be defined as $P_2=(K_2)^{-1}[R_1^2\,|\,-\boldsymbol{t}]$. Therefore, the depth of the point $\boldsymbol{X}$, corresponding to matches $\boldsymbol{x}'$ on the second image and $\boldsymbol{x}$ on the first image, can be calculated as

$$\begin{aligned}depth(\boldsymbol{X};P_1)&=\frac{sign(\det\,(M_1))w_1}{\left\|\boldsymbol{m}_1^3\right\|}\\ depth(\boldsymbol{X};P_2)&=\frac{sign(\det\,(M_2))w_2}{\left\|\boldsymbol{m}_2^3\right\|}\end{aligned} \tag{7}$$

where M_1 and M_2 are the left hand 3 × 3 blocks of P_1 and P_2, respectively. $\boldsymbol{m}_1^3$ and $\boldsymbol{m}_2^3$ are the third rows of M_1 and M_2 respectively, $P_1X = w_1\boldsymbol{x}$ and $P_2X = w_2\boldsymbol{x}'$.

If $depth(\boldsymbol{X};P_1)>0$ and $depth(\boldsymbol{X};P_2)>0$, then the point is in front of both cameras. Therefore, one can test all the four possible choices of P_2 (see Equation (6)), and determine the one which results in positive depths for all the corresponding points. Note that due to high noise, this condition might be incorrect for few points. These points can be recognized and eliminated.

Once the initial linear solution to the rotation matrix and translation vector is determined, the co-planarity equation is used to refine the relative orientation parameters. Suppose that the relative translation vector (air-baseline) is determined as $\boldsymbol{b}=\left(R_1^2\right)^T\boldsymbol{t}$. Therefore, the co-planarity equation can be defined as follows.

$$\det\begin{bmatrix} b_x & b_x + r_{11}.x_{p_2} + r_{12}.y_{p_2} - r_{13}.f & x_{p_1} \\ b_y & b_y + r_{21}.x_{p_2} + r_{22}.y_{p_2} - r_{23}.f & y_{p_1} \\ b_z & b_z + r_{31}.x_{p_2} + r_{32}.y_{p_2} - r_{33}.f & -f \end{bmatrix} = 0 \tag{8}$$

where (x', y') and (x, y) are the image coordinates of matches $\boldsymbol{x}'$ on the second image and $\boldsymbol{x}$ on the first image. Variable f is the camera focal length. The parameters r_{ij} are elements of the rotation matrix $\left(R_1^2\right)^T$

In some techniques, further refinement may also be applied by considering image triplets rather than two-view stereo images. Suppose that there are three images 1, 2, 3 in a triplet. The relative orientation of image 3 with respect to image 1 that are determined from their essential matrix can be refined from the RO parameters of image 3 with respect to image 2 and image 2 with respect to image 1 (He and Habib, 2014). Another way is the robust estimation of the trifocal tensor of all the three views. The trifocal tensor is similar to the fundamental matrix for three images instead of two (Ressl, 2003).

AII.2. Establishing the Local Coordinate System and Incremental Estimation of EO Parameters and 3D Local Coordinates of Corresponding Points

One image in the dataset that is overlapped with the most number of images is selected as the first reference image, and its coordinate system is determined as the local coordinate system (n). Therefore, the camera projection matrix of the first reference image is $P_r = (K_r)^{-1}[I_3 \mid 0]$. The first reference image also decides an arbitrary scale factor for its baseline with one of the images in the dataset. This image is the one that is overlapped with the first reference image (with a large number of tie points) and is also overlapped with many other images in the dataset. Given a target un-orientated image, u, and a reference image, r, the exterior orientation parameters of the target image ($R_u^n, \boldsymbol{t}_u$) are determined as

$$\begin{aligned} R_u^n &= \left(R_r^u \left(R_r^n \right)^T \right)^T \\ \boldsymbol{t}_u &= \boldsymbol{t}_r + \lambda R_u^n \boldsymbol{t} \end{aligned} \tag{9}$$

where R_r^u and $\boldsymbol{t}$ are the relative orientation parameters between the reference and the target image. The parameter λ is the scale factor that is set to an arbitrary value for the first stereo pair and is determined for other pairs using the previously reconstructed points. Given the EO parameters, the local 3D coordinates of the corresponding points between image r and image u can be reconstructed using the following co-linearity equations.

$$\begin{aligned} (K_r)^{-1}\left[R_n^r \mid -R_n^r \boldsymbol{t}_r \right]\left[X^n, Y^n, Z^n, 1 \right]^T &= w_r \left[x_r, y_r, 1 \right] \\ (K_u)^{-1}\left[R_n^u \mid -R_n^u \boldsymbol{t}_u \right]\left[X^n, Y^n, Z^n, 1 \right]^T &= w_u \left[x_u, y_u, 1 \right] \end{aligned} \tag{10}$$

Successively, all the un-oriented images of the dataset should be treated and their EO parameters should be determined. To decide the order of this successive process, at each step, the 3D cloud of tie points

with already oriented images is formed. In this thesis, bundle adjustment is also regularly applied on the already orientated images to avoid error accumulation. Then, for each un-oriented image (image *u*) and any of the already-oriented images (image *r*), the portion of the 3D point cloud that has common tie points with both images is considered (Suppose that this is called the partial point cloud of image *r* and image *u*). The next image to be oriented is the one that its partial point cloud has numerous points, and that the points of its partial cloud are homogeneously spread. To determine the homogeneity, the method of Pierrot and Clery (2011) is used that is based on the inertia matrix of the point cloud. First, the center of mass of the partial point cloud is calculated (simply by averaging the positions of points along each direction). Then, the origin of the partial point cloud is translated to the center of mass, and the inertia matrix of the point cloud along the camera axes passing through the mass center is calculated. The elements of the inertia matrix are the moments of inertia of the point cloud about an axis, which are the measure of the distribution of the point cloud about that axis. The smaller the inertia, the more the mass is concentrated about the axis. Therefore, the smallest value amongst the moments of inertia shows the highest concentration of the point cloud along one axis. Thus, the image whose partial point cloud has the largest smallest moment of inertia has the most homogenous distribution.

AII.3. Geo-referencing the Initial Estimations

This step is optional and required only if the final mapping is to be done in a global coordinate system; i.e. geo-referencing is required. Geo-referencing is done either directly by relying on the EO parameters directly measured via the navigation system, or indirectly by relying on the 3D coordinates of ground control points measured in a global coordinate system. Anyhow, the task of geo-referencing is transforming the 3D object coordinates as well as EO parameters from the local coordinate system (*n*) to the global coordinate system (*e*). Note that this transformation can also be done once the bundle adjustment is performed in a local coordinate system. However, the results would be more accurate if the geo-referencing is repeated inside the bundle adjustment by integrating the global observations into it. Therefore, the best practice is to estimate global values of 3D object coordinates and EO parameters first, and use them as initial values in the bundle adjustment.

A seven parameter Helmert transformation (three translations, three rotations, and one scale) is one of the transformations that can be used to describe the relationship between the local and global coordinate systems. Suppose that (X^e, Y^e, Z^e) are the observations of a ground control point in the global coordinate system, and (X^n, Y^n, Z^n) are the local coordinates of the same point computed via Equation (10). The Helmert transformation can be expressed as Equation (11) (Watson, 2006). Given at least three control points, all the 7 parameters of the transformation ($\boldsymbol{c}$, $\boldsymbol{\varphi}$, μ) can be determined via least squares adjustment.

$$\begin{bmatrix} X^e \\ Y^e \\ Z^e \end{bmatrix} = \mu R(\boldsymbol{\varphi}) \begin{bmatrix} X^n \\ Y^n \\ Z^n \end{bmatrix} + \boldsymbol{c}$$

where

$R(\boldsymbol{\varphi}) = R_1(\alpha)R_2(\beta)R_3(\gamma)$, with $\boldsymbol{\varphi} = (\alpha, \beta, \gamma)^T$ where

$$R_1(\alpha) = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

$$R_2(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$$

$$R_3(\gamma) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{bmatrix}$$

Appendix III. Camera System Reduction in BBA with GHM

Suppose that $\boldsymbol{p}$ is the vector of unknowns including the following elements:

Camera calibration parameters ($\boldsymbol{c}$): u_c unknowns
external orientation paramateres of u_a images in the dataset ($\boldsymbol{a}$): $6u_a$ unknowns
Ground coordinates of u_b tie points ($\boldsymbol{b}$): $3u_b$ unknowns

The vector of observations is $\boldsymbol{m}$ including 2×1 blocks of $\boldsymbol{m}_{\text{ij}}$ as the non-homogenous image coordinate of the it'th point on the j'th image.

General equations with the Gauss-Helmert model would be of form $f(\boldsymbol{p},\boldsymbol{m})=0$ for observation equations and $g(\boldsymbol{p})=0$ for additional constraints on unknowns. Linearizing the equations with Taylor series expansion results in the following equations.

$$\begin{aligned} &f(\boldsymbol{p}^0,\boldsymbol{m}^0)+\left.\frac{\partial f}{\partial \boldsymbol{p}}\right|_{\boldsymbol{p}^0,\boldsymbol{m}^0}\delta\boldsymbol{p}+\left.\frac{\partial f}{\partial \boldsymbol{m}}\right|_{\boldsymbol{p}^0,\boldsymbol{m}^0}\boldsymbol{v}=0 \\ &g(\boldsymbol{p}^0)+\left.\frac{\partial g}{\partial \boldsymbol{p}}\right|_{\boldsymbol{p}^0}\delta\boldsymbol{p}=0 \end{aligned} \tag{1}$$

Renaming variables would result in Equation (2).

$$\begin{aligned} &W_1+A_1\delta\boldsymbol{p}+B\boldsymbol{v}=0 \\ &W_2+A_2\delta\boldsymbol{p}=0 \end{aligned} \tag{2}$$

Suppose that the weight matrix of observations is Q. Suppose that unknowns have a priori weight Q_p as well (if not known, the weight of an unknown is set to 0). Using Lagrange multipliers K_1 and K_2, the objective function to minimize the sum of squared residuals would be as follows.

$$\phi=\boldsymbol{v}^TQ\boldsymbol{v}+\delta\boldsymbol{p}^TQ_c\delta\boldsymbol{p}+2K_1^T(W_1+A_1\delta\boldsymbol{p}+B\boldsymbol{v})+2K_2^T(W_2+A_2\delta\boldsymbol{p}) \tag{3}$$

To minimize ϕ, the derivatives with respect to unknowns should be set to zero as follows.

$$\begin{aligned} &\frac{\partial\phi}{\partial\boldsymbol{v}}=Q\boldsymbol{v}+B^TK_1=0 \\ &\frac{\partial\phi}{\partial X}=Q_c\delta\boldsymbol{p}+A_1^TK_1+A_2^TK_2=0 \\ &\frac{\partial\phi}{\partial K_1}=W_1+A_1\delta\boldsymbol{p}+B\boldsymbol{v}=0 \\ &\frac{\partial\phi}{\partial K_2}=W_2+\delta\boldsymbol{p}=0 \end{aligned} \tag{4}$$

This would result in the normal equation as follows.

$$\begin{bmatrix} Q_x + A_1^T (BQ^{-1}B^T) A_1 & A_2^T \\ A_2 & 0 \end{bmatrix} \begin{bmatrix} \delta \hat{\boldsymbol{p}} \\ \hat{K}_2 \end{bmatrix} = \begin{bmatrix} -A_1^T (BQ^{-1}B^T) W_1 \\ -W_2 \end{bmatrix} \tag{5}$$

Rename the following variables as $H = P_x + A_1^T (BP_v^{-1}B^T) A_1$, $N = -A_1^T (BQ^{-1}B^T) W_1 = \left[N^c, N^a, N^b \right]^T$, $A_2 = \begin{bmatrix} 0 & M^a & M^b \end{bmatrix}$.

Due to the format of its sub-blocks, the final form of the normal equations would be as follows.

$$\begin{bmatrix} H^c & H^{ca} & H^{cb} & 0 \\ H^{ac} & H^a & H^{ab} & \left(M^a\right)^T \\ H^{bc} & H^{ba} & H^b & \left(M^b\right)^T \\ 0 & M^a & M^b & 0 \end{bmatrix} \begin{bmatrix} \delta c \\ \delta a \\ \delta b \\ K_2 \end{bmatrix} = \begin{bmatrix} N^c \\ N^a \\ N^b \\ -W_2 \end{bmatrix} \tag{6}$$

The objective is to solve this system without calculating any large matrix multiplication or inverting and to avoid storing the complete elements of matrices since many of them are zeros.

From the third row of Equation (6), it can be concluded that:

$$\delta b = \left(H^b\right)^{-1} \left(N^b - H^{bc} \delta C - H^{ba} \delta a - \left(M^b\right)^T K_2 \right) \tag{7}$$

Putting δb back to the first row of Equation (6) results in the following equation.

$$\delta c = \left(H^c - H^{cb} \left(H^b\right)^{-1} H^{bc} \right)^{-1} \begin{pmatrix} \left(H^{cb} \left(H^b\right)^{-1} H^{ba} - H^{ca} \right) \delta a \\ + \left(H^{cb} \left(H^b\right)^{-1} \left(M^b\right)^T \right) K_2 \\ + \left(N^c - H^{cb} \left(H^b\right)^{-1} N^b \right) \end{pmatrix} \tag{8}$$

For the sake of representation the components of Equation (8) are renamed as follows, leaving $\delta c = K^c \left(K^a \delta a + K^K K_2 + K^N \right)$.

$$\begin{aligned} K^c &= \left(H^c - H^{cb} \left(H^b\right)^{-1} H^{bc} \right)^{-1} \\ K^O &= H^{cb} \left(H^b\right)^{-1} H^{ba} - H^{ca} \\ K^K &= H^{cb} \left(H^b\right)^{-1} \left(M^b\right)^T \\ K^N &= N^c - H^{cb} \left(H^b\right)^{-1} N^b \end{aligned} \tag{9}$$

Putting the equation of δc back to the third row of Equation (6) gives the following equation for δb which is no more dependant to δc.

$$\delta b = \begin{pmatrix} \left(\left(H^b\right)^{-1} N^b - \left(H^b\right)^{-1} H^{bc} K^c K^N\right) \\ + \left(-\left(H^b\right)^{-1} H^{bc} K^c K^a - \left(H^b\right)^{-1} H^{ba}\right)\delta a \\ + \left(-\left(H^b\right)^{-1} H^{bc} K^c K^K - \left(H^b\right)^{-1} \left(M^b\right)^T\right) K_2 \end{pmatrix} \tag{10}$$

For simplicity, the components of Equation (10) are re-named as follows, leaving $\delta b = U^N + U^a \delta a + U^K K_2$.

$$\begin{aligned} U^N &= \left(\left(H^b\right)^{-1} N^b - \left(H^b\right)^{-1} H^{bc} K^c K^N\right) \\ U^a &= \left(-\left(H^b\right)^{-1} H^{bc} K^c K^a - \left(H^b\right)^{-1} H^{ba}\right) \\ U^K &= \left(-\left(H^b\right)^{-1} H^{bc} K^c K^K - \left(H^b\right)^{-1} \left(M^b\right)^T\right) \end{aligned} \tag{11}$$

Multiplying the left side of the above equation to M^b would give $M^b \delta b$, and from the fourth row of the Equation (6), it can be seen that

$$M^b \delta b = -W_2 - M^a \delta a \,. \tag{12}$$

Therefore, it can be concluded that $-W_2 - M^a \delta a = M^b U^N + M^b U^a \delta a + M^b U^K K_2$. From Equation (12), K_2 can be computed as follows.

$$K_2 = -\left(M^b U^K\right)^{-1} \begin{pmatrix} \left(M^b U^a + M^a\right)\delta a \\ + \left(M^b U^N + W_2\right) \end{pmatrix} \tag{13}$$

For simplicity, the components of Equation (13) are re-named as follows leaving $K_2 = S^N + S^a \delta a$.

$$\begin{aligned} S^N &= -\left(M^b U^K\right)^{-1} \left(M^b U^N + W_2\right) \\ S^a &= -\left(M^b U^K\right)^{-1} \left(M^b U^a + M^a\right) \end{aligned} \tag{14}$$

From the second row of Equation (6), it can be seen that

$$\delta a = (H^a)^{-1} (N^a - H^{ac} \delta c - H^{ab} \delta b - (M^a)^T K_2) \,. \tag{15}$$

Substituting the final equations of other elements (Equation (13), Equation (10) and Equation (8)) to Equation (15) gives the reduced linear system of cameras as follows.

$$\begin{aligned} &\left(I_{u_a} + \left(H^a\right)^{-1} \left(H^{ac} K^c \left(K^a + K^K S^a\right) + H^{ab} \left(U^a + U^K S^a\right) + \left(M^a\right)^T S^a\right)\right)\delta a = \\ &\left(H^a\right)^{-1} \left(N^a - H^{ac} K^c \left(K^N + K^K S^N\right) - H^{ab} \left(U^N + U^K S^N\right) - \left(M^a\right)^T S^N\right) \end{aligned} \tag{16}$$

Appendix IV. Intrinsic Calibration Model and Co-linearity Observation Equations in Gauss-Markov Model

In the Gauss-Markov model, the observations (image points coordinates) are expressed by a non-linear functional relation of parameters. The function is itself linear with respect to the image observations. In the experiments of Section 5.3.2, the following observation equations were defined for GMM,

$$\begin{aligned}\frac{x^p}{-f} &= x^h(1+R_1r^2+R_2r^4+R_3r^6)+2T_1x^hy^h+T_2(r^2+2(x^h)^2)\\ \frac{y^p}{-f} &= y^h(1+R_1r^2+R_2r^4+R_3r^6)+2T_2x^hy^h+T_1(r^2+2(y^h)^2)\end{aligned} \tag{1}$$

where (x^p, y^p) are the coordinates of a feature on an image in its principal coordinate system, (X,Y,Z) are the 3D coordinates of that feature, $(\kappa, \varphi, \omega, X_o, Y_o, Z_o)$ are the EO parameters of the camera, f is the camera focal length, (R_1,R_2,R_3) are radial lens distortion parameters, (T_1,T_2) are tangential lens distortion parameters, and $r^2 = (x^h)^2 + (y^h)^2$.

Coordinates (x^p, y^p) can be determined from the position of the feature in the sensor coordinates system as (u,v) using the pixel size s and position of the principal point (C_x,C_y).

$$\begin{aligned}x^p &= (u+0.5)s - C_x\\ y^p &= -(v+0.5)s + C_y\end{aligned} \tag{2}$$

In Equation (1), (x^h, y^h) are defined as follows,

$$x^h = \frac{X_1^P}{X_3^P}, \quad y^h = \frac{X_2^P}{X_3^P} \tag{3}$$

where

$$\begin{bmatrix}X_1^P\\ X_2^P\\ X_3^P\end{bmatrix} = R(\kappa,\varphi,\omega)\left(\begin{bmatrix}X\\ Y\\ Z\end{bmatrix} - \begin{bmatrix}X_o\\ Y_o\\ Z_o\end{bmatrix}\right). \tag{4}$$

Appendix V. Detailed Tables of Comparative Results for Dense Matching

The following tables are supplementary material to Section 6.6. In these tables, 1-way error percentage, 2-way error percentage, and density of the generated disparity maps are compared. Note that the results of two techniques, RSGM and SURE, are missing from Table AV.1, since their open-access libraries do not provide raw, one-way disparity maps. In addition, in all the tables, for some datasets, the technique ADGC did not generate any result.

Table AV.1. One-way error of disparity maps generated by different techniques of dense matching

Dataset	ICBP	ICBP _Noocc	ICSG	ICSG _Noocc	ICWTA	ICWTA _Noocc	CenWTA	ADGC	ADHierBP
H_Aloe	6.0	7.6	6.4	7.7	22.1	24.8	24.7	21.0	21.9
H_Baby3	17.4	18.2	12.0	14.4	41.9	44.0	44.6	32.5	34.6
H_Cloth4	2.7	5.0	2.5	3.9	20.5	21.6	22.5	19.0	14.1
H_Midd2	43.6	45.4	42.5	43.5	60.2	62.7	61.8		58.5
H_Rocks2	2.6	4.0	3.4	4.6	16.8	18.6	19.1	26.0	19.9
H_Wood1	8.7	11.3	6.9	8.9	41.5	43.6	43.9		32.0
L_Baby1	5.3	6.3	4.5	6.4	22.3	23.6	26.8	19.9	18.2
L_Bowling2	9.8	11.3	9.3	10.7	29.0	31.1	32.8	23.1	37.3
L_flowerpots	11.1	12.9	15.6	18.2	33.2	35.7	36.8	24.7	35.6
L_midd1	37.9	39.9	36.0	37.2	51.5	53.7	52.7	39.8	46.0
L_Monopoly	23.5	25.1	18.9	21.1	42.6	45.4	48.0	66.1	55.5
L_Wood2	8.0	8.8	3.6	5.2	35.2	37.3	39.2	21.2	39.9
H_Art	16.5	18.7	14.0	15.6	42.8	44.7	44.7	30.8	34.9
H_Books	34.4	33.2	28.7	30.0	56.9	59.6	58.8	47.3	36.3
H_Dolls	14.7	16.1	14.7	16.8	42.6	43.6	43.6	30.9	30.2
H_Cones	18.3	19.6	12.2	13.4	48.4	50.2	51.0	63.5	39.8
H_Loundry	36.7	38.2	29.5	31.4	61.0	62.4	61.9	42.2	43.5
H_Moeibuis	20.5	21.8	18.2	19.9	43.5	45.3	44.8	32.0	27.7
H_Reindeer	11.1	12.7	8.9	11.1	43.2	44.6	44.5	28.3	25.7
L_Books	13.2	14.7	9.8	11.1	36.1	37.6	40.3	32.3	46.7
L_Teddy	8.5	10.5	8.3	11.0	26.5	27.6	31.6	29.8	49.4
H_Adirondack	49.7	51.3	44.4	46.2	67.8	70.1	70.5	83.3	81.9
H_Motorcycle	24.9	26.7	20.4	22.7	54.5	55.8	55.7	44.2	43.8
H_Pipes	18.2	19.6	15.0	16.7	41.4	43.2	44.3	36.0	32.4
H_Shelves	51.3	53.1	47.7	49.6	66.4	68.7	68.4	62.5	62.5
L_Adirondack	15.1	16.8	11.8	14.1	38.0	39.6	41.3	60.8	74.5
L_Jadeplant	24.0	25.3	23.6	25.0	38.6	40.5	41.0		39.0
L_MotorE	8.3	9.9	7.2	8.5	26.6	27.9	30.7		84.5
L_Piano	18.9	20.6	17.4	19.5	36.0	37.3	39.1		49.4
L_Playroom	21.8	23.9	18.6	20.5	38.0	40.8	42.3		62.4
L_Playtable	15.3	17.4	16.2	18.0	38.8	39.7	42.4		46.7
L_Vintage	34.5	36.9	33.0	35.2	54.6	57.4	56.8		44.0

Table AV.2. Two-way error of disparity maps generated by different techniques of dense matching

Dataset	ICBP	ICBP _Noocc	ICSG	ICSG _Noocc	ICWTA	ICWTA _Noocc	CenWTA	RSGM	ADGC	AD HierBP	SURE
H_Aloe	2.6	4.0	3.5	6.1	11.0	12.2	13.6	7.5	10.8	11.7	3.4
H_Baby3	6.8	9.1	4.7	6.4	25.3	27.0	28.8	9.9	21.3	20.5	4.1
H_Cloth4	0.8	2.1	1.4	2.8	9.0	10.7	13.1	2.5	9.9	4.3	1.3
H_Midd2	13.9	15.9	7.4	9.1	35.1	38.0	38.0	45.8		41.1	21.8
H_Rocks2	1.2	3.7	2.6	5.3	7.4	9.7	10.2	2.4	10.4	6.2	2.0
H_Wood1	3.2	5.0	2.4	4.7	22.3	23.7	24.8	4.8		16.0	2.1
L_Baby1	2.0	4.3	2.2	3.5	8.6	10.2	15.6	4.5	10.9	10.8	3.1
L_Bowling2	5.6	6.8	6.2	8.3	17.0	18.6	23.6	9.4	19.2	35.1	5.1
L_flowerpots	6.5	8.7	12.4	15.9	22.1	24.8	27.9	15.8	27.4	32.7	6.2
L_midd1	20.3	21.9	19.0	20.4	27.5	29.6	30.8	46.5	32.1	41.4	22.6
L_Monopoly	11.5	13.8	9.2	11.5	23.2	26.2	30.0	25.5	49.1	30.9	12.5
L_Wood2	2.5	5.3	1.1	3.2	16.1	17.6	23.5	5.8	9.7	27.4	5.3
H_Art	8.1	9.2	7.2	8.5	29.9	32.5	32.2	14.3	20.7	21.0	9.2
H_Books	16.9	19.8	12.3	13.8	37.7	40.3	39.5	25.3	30.1	19.6	13.2
H_Dolls	7.7	9.2	9.5	11.2	28.1	31.0	29.7	15.3	22.5	20.4	7.2
H_Cones	6.5	8.9	5.1	7.1	30.2	32.3	32.0	9.0	53.4	23.7	4.0
H_Loundry	19.3	21.5	11.7	14.2	46.3	47.5	48.3	25.5	27.9	24.7	24.6
H_Moeibuis	8.4	9.6	6.7	8.5	23.6	25.6	25.4	15.6	15.5	10.9	7.9
H_Reindeer	4.9	7.9	5.0	7.2	29.5	32.2	30.5	10.2	17.2	12.3	5.2
L_Books	6.5	8.1	6.8	8.4	18.9	21.8	26.5	10.8	20.6	34.3	6.9
L_Teddy	4.7	7.6	4.7	6.0	12.3	14.4	19.4	7.9	15.8	32.8	5.5
H_Adirondack	24.7	26.8	16.1	19.1	48.9	50.5	51.1	37.7	72.7	64.9	16.0
L_Motorcycle	10.7	12.6	8.5	11.1	38.2	40.5	39.8	18.7	25.5	26.8	11.9
H_Pipes	7.4	8.8	5.5	7.6	26.3	28.1	28.7	15.1	21.3	18.5	8.1
H_Shelves	27.8	29.0	17.3	20.3	49.6	52.0	50.8	55.1	50.4	47.6	30.3
L_Adirondack	6.2	7.4	5.8	8.3	18.3	20.6	23.6	11.2	53.8	59.1	5.0
L_Jadeplant	12.5	15.0	11.7	13.7	23.8	26.1	29.1	24.4		17.8	12.9
L_MotorE	4.3	6.8	4.4	5.5	12.9	15.5	18.8	8.2		69.5	5.6
L_Piano	9.7	10.7	10.1	13.1	18.2	19.3	22.3	21.2		35.6	10.0
L_Playroom	13.3	14.9	11.4	13.5	22.7	24.4	26.9	23.1		43.5	11.4
L_Playtable	8.1	10.2	10.7	12.8	20.5	23.4	26.2	17.4		33.6	9.0
L_Vintage	19.0	20.9	15.8	17.5	32.2	33.4	37.5	36.3		22.2	17.8

Table AV.3. Density of disparity maps generated by different techniques of dense matching

Dataset	ICBP	ICBP _Noocc	ICSG	ICSG _Noocc	ICWTA	ICWTA _Noocc	CenWTA	RSGM	ADGC	AD HierBP	SURE
H_Aloe	96	97	96	96	77	77	78	119	82	87	94

Dataset	ICBP	ICBP _Noocc	ICSG	ICSG _Noocc	ICWTA	ICWTA _Noocc	CenWTA	RSGM	ADGC	AD HierBP	SURE
H_Baby3	81	82	87	87	56	56	58	125	72	77	98
H_Cloth4	96	96	97	97	75	75	76	120	80	84	96
H_Midd2	53	53	49	49	37	37	39	121		56	63
H_Rocks2	99	99	99	99	80	80	82	118	74	80	97
H_Wood1	88	89	90	90	56	56	58	117		72	93
L_Baby1	93	93	94	94	73	73	78	112	81	83	95
L_Bowling2	92	92	94	94	70	70	74	123	85	79	98
L_flowerpots	94	95	93	93	68	68	73	133	98	81	100
L_midd1	69	69	70	70	51	51	55	120	68	78	75
L_Monopoly	79	79	85	85	57	57	61	112	40	45	78
L_Wood2	88	89	94	94	61	61	66	117	75	66	87
H_Art	83	83	84	84	56	57	57	132	70	69	86
H_Books	65	65	68	68	45	45	45	114	54	67	85
H_Dolls	87	87	88	88	60	60	60	118	77	79	92
H_Cones	82	82	89	89	53	53	53	118	42	63	97
H_Loundry	62	62	66	65	42	42	42	122	61	61	86
H_Moeibuis	80	80	80	80	56	56	56	117	68	77	80
H_Reindeer	88	88	91	91	54	54	55	123	71	78	91
L_Books	85	86	91	91	61	62	66	114	71	83	93
L_Teddy	93	94	94	94	73	73	77	115	71	58	96
H_Adirondack	53	54	52	52	40	40	40	109	23	30	70
H_Motorcycle	77	77	80	80	50	50	50	116	62	69	88
H_Pipes	83	83	84	84	59	59	60	126	65	76	89
H_Shelves	56	56	53	53	40	40	40	120	49	60	64
L_Adirondack	84	84	89	89	60	61	63	110	68	34	93
L_Jadeplant	77	77	73	74	60	60	65	133		58	79
L_MotorE	96	96	98	98	74	74	77	115		22	98
L_Piano	87	88	91	91	68	68	71	116		70	94
L_Playroom	85	85	88	88	65	65	69	124		52	89
L_Playtable	88	88	90	90	62	62	65	116		69	92
L_Vintage	66	67	62	62	44	44	47	112		63	72